PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/04, 21/02, C12N 15/11
C12N 15/00

A1

(11) International Publication Number: WO 93/00353

(43) International Publication Date: 7 January 1993 (07.01.93)

(21) International Application Number: PCT/US92/05222

(22) International Filing Date: 19 June 1992 (19.06.92)

(30) Priority data:
716,831 20 June 1991 (20.06.91) US
837,195 12 February 1992 (12.02.92) US

(71) Applicant: THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Washington, DC 20000 (US).

(72) Inventors: VENTER, J., Craig ; 1718 Nordic Hill Circle, Silver Spring, MD 20906 (US). ADAMS, Mark, D. ; 12812 Sage Terrace, Germantown, MD 20874 (US).

(74) Agents: SIMPSON, Andrew, H. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION PRODUCT

(57) Abstract

Partial and complete human cDNA and genomic sequences corresponding to particular expressed sequence tags (ESTs). The ESTs are cDNA sequences that are generally between 150 and 500 base pairs in length, are derived from human brain cDNA libraries, correspond to genes transcribed in human brain, and have base sequences identified herein as SEQ ID NOS: 1-315.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION PRODUCT

Technical Field

The present invention relates to newly identified polynucleotide sequences corresponding to transcription products of human genes, and to complete gene sequences associated therewith.

Background

This invention relates to human genes. Identification and sequencing of human genes is a major goal of modern scientific research. The sequence of human genes is more than just a scientific curiosity. For example, by identifying genes and determining their sequences, scientists have been able to make large quantities of valuable human "gene products." These include human insulin, interferon, Factor VIII, tumor necrosis factor, human growth hormone, tissue plasminogen activator, and numerous other compounds. Additionally, knowledge of gene sequences can provide the key to treatment or cure of genetic diseases (such as muscular dystrophy and cystic fibrosis). The present invention represents a quantum leap forward in mankind's knowledge of human gene sequences.

There are several basic concepts of molecular biology which figure prominently in the invention. A brief explanation of those concepts follows. Additional background information and definitions for scientific terms can be found in the literature. See, for example, "Glossary of Genetics, Classical and Molecular" by R. Rieger, A. Michaelis, and M.M. Green (Fifth Edition, Springer-Verlag, New York (1991)). The contents of this and other publications cited in the specification are incorporated by reference herein.

At an initial level, the present invention is based on identification and characterization of gene segments. Genes are the basic units of inheritance. Each gene is a string of connected bases called nucleotides. Most genes are formed of

deoxyribonucleic acid, DNA. (Some viruses contain genes of ribonucleic acid, RNA.) The genetic information resides in the particular sequence in which the bases are arranged. A short sequence of nucleotides is often called a polynucleotide or an oligonucleotide.

Like genes, polypeptides are built from long strings of individual units. These units are amino acids. The nucleotide sequence of a gene tells the cell the sequence in which to arrange the amino acids to make the polypeptide encoded by that gene. In general, chains of up to about 200 amino acids are called polypeptides, while proteins are larger molecules made up of polypeptide subunits; both types of molecules are referred to generally herein as polypeptides. A triplet of nucleotides (codon) in DNA codes for each amino acid or signals the beginning or end of the message (anticodon). The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the original DNA sequence is transcribed.

Generally, enzymes in the cell transcribe the permanent DNA of the gene into a temporary RNA copy, called messenger RNA or mRNA. The mRNA, in turn, can be translated into a polypeptide by the cell. This entire process is called gene expression, and the polypeptide is the gene product encoded by the gene.

Scientists have previously discovered how to reverse the transcription process and copy mRNA back into DNA using an enzyme called reverse transcriptase. The resulting is called complementary DNA, or cDNA. This is schematically shown in the single Figure. When substantially all of the mRNA from one cell or tissue is converted to cDNA at once and cloned into multiple copies of a recombinant vector to allow replication and manipulation in the laboratory, the result is called a cDNA library.

The various types of genes include those which code for polypeptides, those which are transcribed into RNA but are not translated into polypeptides, and those whose functional

significance does not demand that they be transcribed at all. Most genes are found on large molecules of DNA located in chromosomes. Double stranded cDNA carries all the information of a gene. Each base of the first strand is joined to a complementary base (hybridized) in the second strand. The linear DNA molecules in chromosomes have thousands of genes distributed along their length. Chromosomes include both coding regions (coding for polypeptides) and noncoding regions; the coding regions represent only about three percent of the total chromosome sequence.

An individual gene has regulatory regions that include a promoter which directs expression of the gene, a coding region which can code for a polypeptide, and a termination signal. The regulatory DNA sequence is usually a noncoding region that determines if, where, when, and at what level a particular gene is expressed.

The coding regions of many genes are discontinuous, with coding sequences (exons) alternating with noncoding regions (introns). The final mRNA copy of the gene does not include these introns (which can be much longer than the coding region itself), although it does contain certain untranslated regions that usually do not code for the polypeptide gene product. Untranslated sequences at the beginning and end of the mRNA are known as 5'- and 3'-untranslated regions, respectively. This nomenclature reflects the orientation of the nucleotide constituents of the mRNA.

A cDNA is a DNA copy of a messenger RNA, which contains all of the exons of a gene. The cDNA can be thought of as having three parts: an untranslated 5' leader, an uninterrupted polypeptide-coding sequence, and a 3' untranslated region. The untranslated leader and trailing sequences are important for initiation of translation, mRNA stability, and other functions. The untranslated leader and trailing sequences are called 5'- and 3'-untranslated sequences, respectively. The 3' untranslated sequence is usually longer than the 5' untranslated leader, and can be longer than the polypeptide-coding sequence. The untranslated

regions typically have many, randomly-distributed stop codons, and do not display the nonrandom base arrangements found in coding sequences. The 5'-untranslated sequence is relatively short, generally between 20 and 200 bases. The 3'-untranslated sequence is often many times longer, up to several thousand bases.

The translated or coding sequence begins with a translational start codon (AUG or GUG) and ends with a translational stop codon (UAA, UGA, or UAG). Generally, translation begins at the first "start" codon on the mRNA and proceeds to the first "stop" codon. Coding sequences can be distinguished by their nonrandom distribution of bases; numerous computer algorithms have been developed to distinguish coding from noncoding regions in this way.

Human DNA differs from person to person. No two persons (except perhaps identical twins) have identical DNA. While the differences, called allelic variations or polymorphisms, are slight on a molecular level, they account for most of the physical and other observable differences between individuals. It has been estimated that approximately 14 million sequence polymorphism differences exist between individuals.

The ability of one strand of DNA to attach or hybridize to a complementary strand has already been exploited for several purposes. For example, small pieces of DNA (15 to 25 base pairs long) can be made which will hybridize to longer strands of DNA which have a complementary sequence. These short "primers" can be selected such that they hybridize to a specific, unique location on the longer strand. Once the primers have hybridized to their target on the DNA, the polymerase chain reaction (PCR) can be employed to generate millions of copies of (or amplify) the particular segment of DNA between the locations to which two primers are bound. Briefly, this technique allows amplification of a DNA region situated between two convergent primers, using oligonucleotide primers that hybridize to opposite strands. Primer extension proceeds inward across the region between the two primers, and the product of DNA synthesis of one primer serves as a

template for the other primer. Repeated cycles of DNA denaturation, annealing of primers, and extension result in an exponential increase in the number of copies of the region bounded by the primers.

Similarly, a labeled segment of single-stranded DNA can be hybridized to a longer DNA sequence, such as a chromosome, to mark a specific location on the longer sequence. Segments of DNA 50 bases long or longer that hybridize to a unique DNA location in the human genome are extremely unlikely to hybridize elsewhere in the human genome.

The Human Genome Project is an effort to sequence all human DNA (the human genome). The human genome is estimated to comprise 50,000 - 100,000 genes, up to 30,000 of which might be expressed in the brain (Sutcliffe, **Ann. Rev. Neurosci. 11**:157 (1988)). Once dedicated human chromosome sequencing begins in three to five years, it was expected that 12-15 years will be required to complete the sequence of the genome (**Report of the Ad Hoc Program Advisory Committee on Complex Genomes**, Reston, Va., Feb. 1988, D. Baltimore Ed. (NIH, Bethesda, Md, 1988)). At that rate, the majority of human genes would remain unknown for at least the next decade. The present invention can greatly accelerate the pace at which human genes can be identified and mapped. Most gene researchers, in conjunction with publication of their results in this field, submit sequence data to the GenBank database. Prior to the present invention, GenBank listed the sequences of only a few thousand human genes and less than two hundred human brain mRNAs (GenBank Release 66.0, December, 1990).

The role of sequencing complementary DNA (cDNA), reverse transcribed from mRNA, as a part of the human genome project has been vigorously debated since the idea of determining the complete nucleotide sequence of humans first surfaced. The coding sequence of all human genes represents most of the information content of the genome, but only 3-5% of the total DNA. In contrast, cDNA (which is only made from the transcription product of active genes, is one-half to three-fourths the remainder being 5'- and 3'-untranslated sequence

meaningful genetic information. Thus, some have argued that cDNA sequencing should take precedence over genomic sequencing (Brenner, **CIBA Found. Symp. 149**:6 (1990)). However, until now, such arguments have not been heeded.

Genomic sequencing proponents have argued the difficulty of finding every mRNA expressed in all tissues, cell types, and developmental states, and that much valuable information from intronic and intergenic regions, including control and regulatory sequences, will be missed by cDNA sequencing. (Report of the Committee on Mapping and Sequencing the Human Genome, National Research Council (**National Academy Press**, Washington, D.C. 1988)). Further, sequencing of transcribed regions of the genome using cDNA libraries has heretofore been considered impractical or unsatisfactory. Libraries of cDNA were believed to be dominated by repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes comprising common or housekeeping sequences. It was believed that cDNA libraries would provide few sequences corresponding to structural and regulatory polypeptides or peptides. See, for example, Putney, et al., **Nature 302**:718-721 (1983). Putney, et al. sequenced over 150 clones from a rabbit muscle cDNA library and identified clones for 13 of the 19 known muscle polypeptides, including one new isotype but no unknown coding sequences.

Another perceived drawback of cDNA sequencing was that some mRNAs are abundant, and some are rare. The cellular quantities of mRNA from various genes can vary by several orders of magnitude. This led critics to believe that most information obtained from cDNA sequencing would be repetitious and useless.

The present invention demonstrates that, despite such skepticism, cDNA sequencing now provides a rapid method for obtaining enormous amounts of valuable genetic information and DNA products of great utility for the biotechnology and pharmaceutical industries. Not only can many distinct cDNAs be isolated and sequenced, even partial cDNAs can be used, with conventional, well-understood methods, to isolate entire

genes, and to determine the chromosomal locations and biological functions of these genes. As is demonstrated here, fragments of only a few hundred bases are sufficient, in many cases, to identify the probable function of a new human gene if it is similar in structure to a gene from another animal, or from plants or bacteria. Similarly, even fragments of untranslated regions of a cDNA can be used to: i) isolate the coding sequence of the cDNA; ii) isolate the complete gene; iii) determine the position of the gene on a human chromosome, and hence the potential of the gene to cause a human genetic disease; and iv) determine the function of the gene by means of experiments in which the function of the native gene is disrupted by the addition of a short DNA fragment to the cell, e.g., using triple helix or antisense probes.

Because coding regions comprise such a small portion of the human genome, identification and mapping of transcribed regions and coding regions of chromosomes is of significant interest. There is a corresponding need for reagents for identifying and marking coding regions and transcribed regions of chromosomes. Furthermore, such human sequences are valuable for chromosome mapping, human identification, identification of tissue type and origin, forensic identification, and locating disease-associated genes (i.e., genes that are associated with an inherited human disease, whether through mutation, deletion, or faulty gene expression) on the chromosome.

SUMMARY OF THE INVENTION

Contrary to the expectations of the scientific community, cDNA screening and sequencing techniques have now been used to discover a large number of heretofore unknown human genes. Disclosed herein are over 300 new human polynucleotide sequences. The novelty of these sequences has been established through comparison to both nucleotide sequence databases and amino acid sequence databases. Surprisingly, approximately 80% of the sequences generated were unrelated to any sequences previously described in the literature.

The sequences of the present invention were ascertained using a fast approach to cDNA characterization. This approach could facilitate the tagging of most expressed human genes within a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, provide new DNA-based therapeutics and diagnostics, and provide other valuable nucleotide reagents.

The sequences disclosed herein, styled Expressed Sequence Tags ("ESTs"), are markers for human genes actually transcribed *in vivo*. Techniques are disclosed for using these ESTs to obtain the full coding region of the corresponding gene. The use of ESTs, complete coding sequences, or fragments thereof for marking chromosomes, for mapping locations of expressed genes on chromosomes, for individual or forensic identification, for mapping locations of disease-associated genes, for identification of tissue type, and for preparation of antisense sequences, probes, and constructs is discussed in detail below. Unlike the random genomic DNA sequence tagged sites (STSs) (Olson et al., **Science 245**:1434 (1989)), ESTs point directly to expressed genes.

Various aspects of the present invention thus include the individual ESTs, corresponding partial and complete cDNA, genomic DNA, mRNA, antisense strands, triple helix probes, PCR primers, coding regions, and constructs. Also, where one skilled in the art is enabled by this specification to prepare expression vectors and polypeptide expression products, they are also within the scope of the present invention, along with antibodies, especially monoclonal antibodies, to such expression products.

BRIEF DESCRIPTION OF THE DRAWING

The single drawing Figure schematically illustrates the progression from chromosome to gene to mRNA to cDNA.

DETAILED DESCRIPTION OF THE INVENTION

The detailed description that follows provides not only the actual sequence of each new EST, but also explains how the ESTs were obtained, how to obtain the corresponding complete cDNA sequence and the corresponding genomic DNA sequence, how to make DNA constructs from the ESTs and corresponding sequences, how to use those sequences as reagents in molecular biology and other fields, how to produce gene products from the ESTs and corresponding sequences and antibodies to those gene products, and the functional categories of many ESTs and corresponding genes. Furthermore, numerous actual working examples and predictive examples are provided to demonstrate and exemplify numerous aspects of the invention.

I. ESTs from cDNA Libraries

The sequences of the present invention were isolated from commercially available and custom made cDNA libraries using a rapid screening and sequencing technique. In general, the method comprises applying conventional automated DNA sequencing technology to screening clones, advantageously randomly selected clones, from a cDNA library. Preferably, the library is initially "enriched" through removal of ribosomal sequences and other common sequences prior to clone selection. According to the present method, ESTs are generated from partial DNA sequencing of the selected clones. The ESTs of the present invention were generated using low redundancy of sequencing, typically a single sequencing reaction. While single sequencing reactions may have an accuracy as low as 97%, this nevertheless provides sufficient fidelity for identification of the sequence and design of PCR primers.

Most human genes can be identified by EST sequencing from libraries of cDNA copies of messenger RNAs. However, some genes are expressed only at specific times during embryonic development, or only in small amounts in a few specific cell types. Other genes have mRNAs that are degraded very quickly by the cell in which they are expressed. If any of these are the case, transcripts of the gene will not be represented in

cDNA libraries so the gene will not be identifiable by EST sequencing. A new method called "exon amplification", however, can be used to isolate and identify transcripts of such genes.

Exon amplification works by artificially expressing part or all of a gene that is contained in a cloned fragment of genomic DNA such as a cosmid or yeast artificial chromosome (YAC). The gene is cloned into a special vector, designed at MIT, that uses control elements from virus genes to express the protein-coding exons of the human gene of interest. Exon trapping shows considerable promise as a general technique for identifying those genes in the human genome that cannot be found by cDNA cloning and EST sequencing. Exon amplification will also be useful for identifying the genes in regions of genomic DNA to which disease genes have been mapped. The exon amplification method can be used directly with the cosmid and YAC clones from human chromosomes that are being obtained by both NIH and DOE supported human genome centers.

ESTs comprise DNA sequences corresponding to a portion of nuclear encoded messenger RNA. An EST is of sufficient length to permit: (1) amplification of the specific sequence from a cDNA library, e.g., by polymerase chain reaction (PCR); (2) use of a synthetic polynucleotide corresponding to a partial or complete sequence of the EST as a hybridization probe of a cDNA library, generally having 30 - 50 base pairs; or (3) unique designation of the pure cDNA clone from which the EST was derived (the EST clone) for use as a hybridization probe of a cDNA library. Preferably, EST-derived primer pairs and sequences amplify or detectably hybridize to a sequence from a genomic library.

It has been found that sufficient information is contained in the 150-400 base ESTs from one sequencing run to effect preliminary identification and exact chromosome mapping. Accordingly, the ESTs disclosed herein are generally at least 150 base pairs in length. The length of an EST is determined by the quality of sequencing data and the length of the cloned cDNA. Raw data from the automated sequencers is

edited to remove low quality sequence at the end of the sequencing run. High quality sequences (usually a result of sequencing templates without excessive salt contamination) generally give about 400 bp of reliable sequence data; other sequences give fewer bases of reliable data. A 150 bp EST is long enough to be translated into a 50 amino acid peptide sequence. This length is sufficient to observe similarities when they exist in a database search. Furthermore, 150 bp is long enough to design PCR primers from each end of the sequence to amplify the complete EST. Sequences shorter than 150 bp are difficult to purify and use following PCR amplification. Furthermore, a 150 bp polynucleotide is likely to give a very strong signal with low background in a screen of a genomic library.

Finally, it is highly unlikely that a sequence of the same 150 bp exists in any genes in the genome besides the one tagged by the EST. Some closely related gene family members have very similar nuclectide sequences, but no examples of pairs of human genes with long segments of identical sequence have been reported to date. For instance, there are three known β-tubulin genes in humans. Several ESTs were found that matched one or another of these tubulin genes, but several new members of this gene family were also found and could be clearly distinguished from the three known members. ESTs that match perfectly to several different genes can be detected by hybridizing to chromosomes: if many chromosomal loci are observed, the sequence (or a close variant) is present in more than one gene. This problem can be circumvented by using the 3'-untranslated part of the cDNA alone as a probe for the chromosomal location or for the full-length cDNA or gene. The 3'-untranslated region is more likely to be unique within gene families, since there is no evolutionary pressure to conserve a coding function of this region of the mRNA.

As demonstrated in the Examples that follow, ESTs can be used to map the expressed sequence to a particular chromosome. In addition, ESTs can be expanded to provide the full coding

regions, as detailed below. In this manner, previously unknown genes can be identified.

While a variety of cDNA libraries can be used to obtain ESTs, human brain cDNA libraries are exemplified and represent a preferred embodiment. Suitable cDNA libraries can be freshly prepared or obtained commercially, e.g., as shown in Examples 1 and 9. The cDNA libraries from the desired tissue are preferably preprocessed by conventional techniques to reduce repeated sequencing of high and intermediate abundance clones and to maximize the chances of finding rare messages from specific cell populations. Preferably, preprocessing includes the use of defined composition prescreening probes, e.g., cDNA corresponding to mitochondria, abundant sequences, ribosomes, actins, myelin basic polypeptides, or any other known high abundance peptide; these prescreening probes used for preprocessing are generally derived from known ESTs. Other useful preprocessing techniques include subtraction, which preferentially reduces the population of certain sequences in the library (e.g., see A. Swaroop et al., **Nucl. Acids Res. 19**:1954 (1991)), and normalization, which results in all sequences being represented in approximately equal proportions in the library (Patanjali et al, **Proc. Natl. Acad. Sci. USA 88**:1943 (1991)).

The cDNA libraries used in the present method will ideally use directional cloning methods so that either the 5' end of the cDNA (likely to contain coding sequence) or the 3' end (likely to be a non-coding sequence) can be selectively obtained.

Libraries of cDNA can also be generated from recombinant expression of genomic DNA. After they are amplified, ESTs can be obtained and sequenced, e.g., as illustrated in Example 9.

The sequences of the present invention include the specific sequences set forth in the Sequence Listing and designated SEQ ID NO: 1 - SEQ ID NO: 315. In one aspect of this embodiment, the invention relates to those sequences of SEQ ID NOS: 1 - 315 that comprise the cDNA coding sequences for polypeptides having less than 95% identity with known

amino acid sequences (see Table 2) and more preferably less than 90% or 85% identity. In a second aspect, the invention relates to those sequences of SEQ ID NOS: 1 - 315 that encode polypeptides having no similarity to known amino acid sequences (see Examples that follow). Precisely because they do not contain coding regions and are therefore more unique in their sequence structures, those sequences which meet neither of the preceding criteria can be most useful and are generally preferred for mapping.

Consistent with the NIH mission and its responsibilities to disseminate knowledge and share the tangible fruits of its research, the present inventors have taken a number of steps to facilitate sequence data and clone availability. All EST sequences have been submitted to GenBank. The corresponding cDNA clones have been submitted to the American Type Culture Collection and information on clones and sequences has been submitted to the Genome Data Base (Pearson, **P. Nucl. Acids Res. 19 (Suppl.)**: 2237-9 (1991)).

II. <u>Complete Coding Sequences from ESTs</u>

The ESTs of the present invention generally represent relatively small coding regions or untranslated regions of human genes. Although most of these sequences do not code for a complete gene product, the ESTs of the present invention are highly specific markers for the corresponding complete coding regions. The ESTs are of sufficient length that they will hybridize, under stringent conditions, only with DNA for that gene to which they correspond. Suitably stringent conditions comprise conditions, for example, where at least 95%, preferably at least 97% or 98% identity (base pairing), is required for hybridization. This property permits use of the EST to isolate the entire coding region and even the entire sequence. Therefore, only routine laboratory work is necessary to parlay the unique EST sequence into the corresponding unique complete gene sequence.

Thus, each of the ESTs of the present invention "corresponds" to a particular unique human gene. Knowledge of

the EST sequence permits routine isolation and sequencing of the complete coding sequence of the corresponding gene. The complete coding sequence is present in a full-length cDNA clone as well as in the gene carried on genomic clones. Therefore, each EST "corresponds" to a cDNA (from which the EST was derived), a complete genomic gene sequence, a polypeptide coding region (which can be obtained either from the cDNA or genomic DNA), and a polypeptide or amino acid sequence encoded by that region.

The first step in determining where an EST is located in the cDNA is to analyze the EST for the presence of coding sequence, e.g., as described in Example 12. The CRM program predicts the extent and orientation of the coding region of a sequence. Based on this information, one can infer the presence of start or stop codons within a sequence and whether the sequence is completely coding or completely non-coding. If start or stop codons are present, then the EST can cover both part of the 5'-untranslated or 3'-untranslated part of the mRNA (respectively) as well as part of the coding sequence. If no coding sequence is present, it is likely that the EST is derived from the 3'-untranslated sequence due to its longer length and the fact that most cDNA library construction methods are biased toward the 3' end of the mRNA.

One general procedure for obtaining complete sequences from ESTs is as follows:

1. Purify selected human DNA from an EST clone (the cDNA clone that was sequenced to give the EST), e.g., by endonuclease digestion using ECOR1, gel electrophoresis, and isolation of the aforementioned clone by removal from low-melting agarose gel.

2. Radiolabel the isolated insert DNA, e.g., with ^{32}P labels, preferably by nick translation or random primer labeling.

3. Use the labeled EST insert as a probe to screen a lambda phage cDNA library or a plasmid cDNA library.

4. Identify colonies containing clones related to the

probe cDNA and purify them by known purification methods.

5. Nucleotide sequence the ends of the newly purified clones to identify full length sequences.

6. Perform complete sequencing of full length clones by Exonuclease III digestion or primer walking. Northern blots of the mRNA from various tissues using at least part of the EST clone as a probe can optionally be performed to check the size of the mRNA against that of the purported full length cDNA.

An EST is a specific tag for a messenger RNA molecule. The complete sequence of that messenger RNA, in the form of cDNA, can be determined using the EST as a probe to identify a cDNA clone corresponding to a full-length transcript, followed by sequencing of that clone. The EST or the full-length cDNA clone can also be used as a probe to identify a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons, and introns.

ESTs are used as probes to identify the cDNA clones from which an EST was derived. ESTs, or portions thereof, can be nick-translated or end-labelled with ^{32}P using polynucleotide kinase and labelling methods known to those with skill in the art (**Basic Methods in Molecular Biology**, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986). The lambda library can be directly screened with the labelled ESTs of interest or the library can be converted en masse to pBluescript (Stratagene, La Jolla, California) to facilitate bacterial colony screening. Both methods are well known in the art.

Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured and the DNA is fixed to the filters. The filters are hybridized with the labelled probe using hybridization conditions described by Davis et al. The ESTs, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are

compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected, expanded and the DNA is isolated from the colonies for further analysis and sequencing.

The ESTs can additionally be used to screen Northern blots of mRNA obtained from various tissues or cell cultures, including the tissue of origin of the EST clone. Northern analysis will most often produce one to several positive bands. The bands can be selected for further study based on the predicted size of the mRNA.

Positive cDNA clones in phage lambda are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the EST and the other primer from the vector. Clones with a larger vector-insert PCR product than the original EST clone are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size on a Northern blot.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined. The preferred method is to use exonuclease III digestion (McCombie, W.R, Kirkness, E., Fleming, J.T., Kerlavage, A.R., Iovannisci, D.M., and Martin-Gallardo, R., **Methods: 3**: 33-40, 1991). A series of deletion clones is generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

A similar screening and clone selection approach can be applied to obtaining cosmid or lambda clones from a genomic DNA library that contains the complete gene from which the EST was derived (Kirkness, E.F., Kusiak, J.W., Menninger, J., Gocayne, J.D., Ward, D.C., and Venter, J.C. **Genomics 10**: 985-995 (1991). Although the process is much more laborious, these genomic clones can also be sequenced in their entirety.

A shotgun approach is preferred to sequencing clones with inserts longer than 10 kb (genomic cosmid and lambda clones).

In shotgun sequencing, the clone is randomly broken into many small pieces, each of which is partially sequenced. The sequence fragments are then aligned to produce the final contiguous sequence with high redundancy. An intermediate approach is to sequence just the promoter region and the intron-exon boundaries and to estimate the size of the introns by restriction endonuclease digestion (ibid.).

Using the sequence information provided herein, the polynucleotides of the present invention can be derived from natural sources or synthesized using known methods. The sequences falling within the scope of the present invention are not limited to the specific sequences described, but include human allelic and species variations thereof and portions thereof of at least 15-18 bases. (Sequences of at least 15-18 bases can be used, for example, as PCR primers or as DNA probes.) In addition, the invention includes the entire coding sequence associated with the specific polynucleotide sequence of bases described in the Sequence Listing, as well as portions of the entire coding sequence of at least 15-18 bases and allelic and species variations thereof. Furthermore, to accommodate codon variability, the invention includes sequences coding for the same amino acid sequences as do the specific sequences disclosed herein. Finally, although the error rate in the automated sequencing used in the present invention is small, there remains some chance of error. Therefore, claims to particular sequences should not be so narrowly construed as to require inclusion of erroneously identified bases or to exclude corrections.

Any specific sequence disclosed herein can be readily screened for errors by resequencing each EST in both directions (i.e., sequence both strands of cDNA).

The sequences, constructs, vectors, clones, and other materials comprising the present invention can advantageously be in enriched or isolated form. As used herein, "enriched" means that the concentration of the material is at least about

2, 5, 10, 100, or 1000 times its natural concentration (for example), advantageously 0.01%, by weight, preferably at least about 0.1% by weight. Enriched preparations of about 0.5%, 1%, 5%, 10%, and 20% by weight are also contemplated. Further, removal of clones corresponding to ribosomal RNA and "housekeeping" genes and clones without human cDNA inserts results in a library that is "enriched" in the desired clones.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

It is also advantageous that the sequences be in purified form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^6-fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

In a cDNA library there are many species of mRNA represented. Each cDNA clone can be interesting in its own right, but must be isolated from the library before further

experimentation can be completed. In order to sequence any specific cDNA, it must be removed and separated (i.e. isolated and purified) from all the other sequences. This can be accomplished by many techniques known to those of skill in the art. These procedures normally involve identification of a bacterial colony containing the cDNA of interest and further amplification of that bacteria. Once a cDNA is separated from the mixed clone library, it can be used as a template for further procedures such as nucleotide sequencing.

Although claims to large numbers of ESTs and corresponding sequences are presented herein, the invention is not limited to these particular groupings of sequences. Thus, individual sequences are considered as applicants' discoveries or inventions, as are subgroupings of sequences. All of the functional subgroupings set forth in the tables define groupings for which separate claims are contemplated as being within the scope of this invention. Moreover, in addition to claims to individual clones, it is intended that the present disclosure also support claims to numerical subgroupings. Thus, subgroupings of 50 ESTs (and corresponding sequences) are contemplated (e.g., SEQ ID NOS 1-50, 51-100, 101-150, etc.) as being within the scope of this invention, as are subgroupings of 5, 10, 25, 100, 200, and 300 ESTs and corresponding sequences.

III. DNA Constructs

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a sense or antisense orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example.

Bacterial: pBs, phagescript, ϕX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia).

Eukaryotic: pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described construct. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a procaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., **Basic Methods in Molecular Biology**, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product coded by the recombinant sequence. Alternatively, the encoded polypeptide can be synthetically produced by conventional peptide synthesizers.

Certain ESTs have already been preliminarily categorized by analogy to related sequences in other organisms (see Table 2). Table 10 of Example 8 categorizes particular ESTs broadly as metabolic, regulatory, and structural sequences where known. Constructs comprising genes or coding sequences corresponding to each of these categories are, therefore, specifically and individually contemplated.

Table 11 more particularly separates 27 new ESTs into 11 categories using a different criteria. These are genes related to cell surface; developmental control; energy metabolism; kinase and phosphatase; oncogenes; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transcription, translation, and subcellular localization; and transcription factors. Table 11 further identifies the EST by the particular gene product for which it apparently codes. Each of these categories individually comprises a preferred category of EST, and preferred constructs and resulting polypeptide can be prepared from those ESTs or the corresponding complete gene sequence.

IV. ESTs and Corresponding Sequences as Reagents

Each of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type. In addition, these sequences can be used as diagnostic probes suitable for use in genetic linkage analysis (polymorphisms). Further, the sequences can be used as probes for locating gene regions associated with genetic disease, as explained in more detail below.

The EST and complete gene sequences of the present invention are also valuable for chromosome identification. Each sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The present invention constitutes a major expansion of available chromosome markers.

Using the techniques described in Example 3 or 4, ESTs and their corresponding complete sequences can be mapped to chromosomes. The mapping of ESTs and cDNAs to chromosomes according to the present invention is an important first step

in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the ESTs. Computer analysis of the ESTs is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the EST will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular EST to a particular chromosome. Three or more clones can be assigned per day using a single thermal cycler. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map an EST to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific cDNA libraries. Results of mapping ESTs to chromosomal segments are listed in Tables 3 and 4.

Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clone from which the EST was derived, and the longer the better. 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., **Human Chromosomes: a Manual of**

Basic Techniques; Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping (see Tables 8 and 9).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, **Mendelian Inheritance in Man** (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from

polymorphisms.

In addition to the foregoing, the sequences of the invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, **Nucl. Acids Res. 6:** 3073 (1979); Cooney et al, **Science 241:** 456 (1988); and Dervan et al, **Science 251:** 1360 (1991)) or to the mRNA itself (antisense - Okano, **J. Neurochem. 56:** 560 (1991); **Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression**, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be efficient in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

The present invention is also a useful tool in gene therapy, which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect. The high specificity of the cDNA probes according to this invention have promise of targeting such gene locations in a highly accurate manner.

The sequences of the present invention, as broadly defined, are also useful for identification of individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current

limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP.

However, RFLP is a pattern based technique, which does not directly focus on the actual DNA sequence of the individual. The sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take an EST of the invention and prepare two PCR primers from the 5' and 3' ends of the EST. These are used to amplify an individual's DNA, corresponding to the EST. The amplified DNA is sequenced.

Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences, due to allelic differences. The sequences of the present invention can be used to particular advantage to obtain such identification sequences from individuals and from tissue, as explained in Examples 10 - 12. The EST sequences from Example 1 and the complete sequences from Example 11 uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the ESTs or complete coding sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of Table 9 for example, could comfortably provide positive individual identification with a panel of perhaps 100 to 1,000 primers

which each yield a noncoding amplified sequence of 100 bp. If predicted coding sequences, such as those from Table 6, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from ESTs or complete sequences of this invention is used to generate a unique ID database for an individual, those same reagents can later be used to identify tissue from that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

Another use for DNA-based identification techniques is in forensic biology. PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc. In one prior art technique, gene sequences are amplified at specific loci known to contain a large number of allelic variations, for example the DQα class II HLA gene (Erlich, H., **PCR Technology**, Freeman and Co. (1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQα class II HLA gene.

The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions (see, e.g., Tables 8 and 9) are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of the present invention. Such reagents can comprise complete ESTs or corresponding coding regions, or fragments of either of at least 15 bp, preferably at least 18 bp.

There is also a need for reagents capable of identifying

the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the ESTs or complete sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue culture for contamination.

V. Production of Polypeptide Corresponding to ESTs

As previously explained, each EST corresponds not only to a coding region, but also to a polypeptide. Once the coding sequence is known, or the gene is cloned which encodes the polypeptide, conventional techniques in molecular biology can be used to obtain the polypeptide.

At the simplest level, the amino acid sequence encoded by the polynucleotide sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. (Fragments are useful, for example, in generating antibodies against the native polypeptide.)

Alternatively, the DNA encoding the desired polypeptide can be inserted into a host organism and expressed. The organism can be a bacterium, yeast, cell line, or multicellular plant or animal. The literature is replete with examples of suitable host organisms and expression techniques. For example, naked polynucleotide (DNA or mRNA) can be injected directly into muscle tissue of mammals, where it is expressed. This methodology can be used to deliver the polypeptide to the animal, or to generate an immune response against a foreign polypeptide (Wolff, et al., **Science 247**:1465 (1990); Felgner, et al., **Nature 349**:351 (1991). Alternatively, the coding sequence, together with appropriate regulatory regions (i.e., a construct), can be inserted into a vector, which is then used to transfect a cell. The cell (which may or may not be part of a larger organism) then expresses the polypeptide. (See Example 13.

Antibodies generated against the polypeptide corresponding to a sequence of the present invention can be obtained by direct injection of the naked polypeptide into an animal (as above) or by administering the polypeptide to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide. Moreover, a panel of such antibodies, specific to a large number of polypeptides, can be used to identify and differentiate such tissue.

VI. Examples

Certain aspects of the present invention are described in greater detail in the non-limiting Examples that follow.

EXAMPLE 1

cDNA Sequences Determined by Random Clone Selection: First set

METHODOLOGY:

With reference to the data presented in Table 1, lambda ZAP libraries were converted en masse to pBluescript plasmids, transfected into E. coli XL1-Blue cells, and plated on X-gal/IPTG/ampicillin plates. A total of 1058 clones were picked at random from three human brain cDNA libraries: fetal brain, two-year-old hippocampus, and two-year-old temporal cortex (Stratagene catalog #936206, 936205, 935, respectively. Stratagene, 11099 N. Torrey Pines Rd., La Jolla, CA 92037). An analysis of these clones is summarized in Table I (see below) In addition, clones selected from the hippocampus library were also analyzed after subtractive hybridization with the fibroblast library. These results are listed in the "Hippocampus Subtracted" column of Table 1. Templates for DNA sequencing were PCR products or plasmids prepared by the alkaline lysis method. About half of the templates prepared by PCR failed to yield an amplified fragment suitable for sequencing. This was primarily due to use of PCR conditions that minimized the need for further purification of the product but also selected against amplification of long inserts (5 µl fresh or frozen overnight culture of E. coli carrying the pBluescript plasmid, 7.5 µM each dNTP, and 0.1 µM each primer for 35 cycles: 94°C, 40 sec; 55°C, 40 sec; 72°C, 90 sec). A further percentage of the PCR-generated templates failed to sequence, largely due to primer-dimer or other amplification artifacts. Qiagen™ columns improved the percentage of plasmid templates, increasing the yields of usable sequence from about 60% with a standard alkaline lysis protocol to over 90%. Overall, 11% PCR-generated templates and 49% plasmid templates resulted in usable sequence. Dideoxy chain termination sequencing reactions were performed with fluorescent dye-labeled M13 universal or reverse primers.

After a cycle sequencing protocol, carried out in a Perkin-Elmer thermal cycler, sequencing reactions were run on an Applied Biosystems, Inc. (Foster City, CA) 373A automated DNA sequencer. (Cycle sequencing was performed in a Perkin Elmer Thermal Cycler for 15 cycles of 95°C, 30 sec; 60°C, 1 sec; 70°C, 60 sec and 15 cycles of 95°C, 30 sec; 70°C, 60 sec with the Applied Biosystems, Inc. Taq Dye Primer Cycle Sequencing Core Kit protocol). Some sequencing reactions were performed on an ABI robotic workstation (Cathcart, **Nature 347**: 310 (1990) hereby incorporated by reference).

RESULTS:

Singe-run DNA sequence data were obtained from 609 randomly chosen cDNA clones. The number of clones sequenced from each library is summarized in Table 1. Double-stranded cDNA clones in the pBluescript vector were sequenced by a cycle sequencing protocol with dye-labeled primers and Applied Biosystems, Inc. 373A DNA Sequences. The average length of usable sequence was 397 bases with a standard deviation of 99 bases.

Subtractive hybridization has been used successfully to reduce the population of highly represented sequences in a cDNA library by selectively removing sequences shared by another library. (Schmid and Girou, **Neurochem. 48**: 307 (1987); Fargnoli et al, **Anal. Biochem. 187**: 364 (1990); Duguid and Dinauer, **Nucl. Acids. Res. 18**: 2789 (1990); Schweinfest, et al, **Genet. Anal. Techn. Appl. 7**: 64 (1990); Travis and Sutcliffe, **Proc. Natl. Acad. Sci. USA 85**: 1696 (1988); Kato, **Eur. J. Neurosci. 2**: 704 (1990)). Subtractive hybridization was therefore tested as a way of enhancing the number of brain-specific clones in the hippocampus library by hybridizing the hippocampus library with a WI38 human lung fibroblast cell line cDNA library and removing the common sequences (Schweinfest et al, **Genet. Anal. Techn. Appl. 7**: 64 (1990); Sive and St. John, **Nucl. Acids Res. 16**: 10937 (1988)). Clones from this subtraction are listed in the column

"Hippocampus Subtracted" in Table 1.

The EST sequences from this Example 1 are identified as SEQ ID NOs 1-315.

TABLE 1. cDNA Library Composition Determined
By Random Clone Sequencing

	cDNA Library							
	Hippocampus		Hippocampus Subtracted		Fetal Brain		Temporal Cortex	
EST Category	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Databases Match--Human	48	12.8	10	8.6	3	7.9	6	7.5
Mitochondrial Genes	39	10.4	14	12.2	6	15.8	0	0
Repeats: Alu, Line-1, etc.	10	2.7	7	6.0	0	0	11	13.8
Ribosomal RNA	32	8.6	7	6.0	4	10.5	0	0
Other Nuclear Genes	32	8.6	7	6.0	5	13.2	4	5.0
Database Match--Other	160	42.8	44	37.9	20	52.6	6	7.5
No Database Match	53	14.1	24	20.7	0	0	27	33.7
poly A Insert	1	0.3	3	2.6	0	0	26	32.5
No Insert								

EXAMPLE 2

EST Characterization: First Set

ESTs including SEQ ID NOs 1-315 were analyzed as follows. Initially, the EST sequences were examined for similarities in the GenBank nucleic acid database (GenBank Release 65.0), Protein Information Resource Release 26.0 (PIR), and ProSite (MacPattern from the EMBL data library, Fuchs R. **Comput. Appl. Biosci.** 7: 105 (1990) Release 5.0 were used). BLAST was used to search Genbank and the PIR (both maintained by the National Center for Biotechnology Information) ESTs without exact GenBank matches were translated in all six reading frames and each translation was compared with the protein sequence database PIR and the ProSite protein motif database. Comparisons with the ProSite motif database were done by means of the program MacPattern from the EMBL Data Library. GenBank and PIR searches were conducted with the "basic local alignment search tool" programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, **J. Mol. Biol. 215**: 403 (1990)). PIR searches were run on the National Center for Biotechnology Information BLAST network service. The BLAST programs contain a very rapid database-searching algorithm that searches for local areas of similarity between two sequences and then extends the alignments on the basis of defined match and mismatch criteria. The algorithm does not consider the potential gaps to improve the alignment, thus sacrificing some sensitivity for a 6-80 fold increase in speed over other database-searching programs such as FASTA (Pearson and Lipman, **Proc. Natl. Acad. Sci. USA**, 85: 2444 (1988)).

Sequence similarities identified by the BLAST programs were considered statistically significant with a Poisson P-value than 0.01. The Poisson P-value less than the probability of as high a score occurring by chance given the number of residues in the query sequence and the database. After the BLASTN search, 31 unmatched ESTs were compared against GenBank by FASTA to determine if significant matches

were missed due to the use of BLASTN for the database search. No additional statistically significant matches were found. Statistical significance does not necessarily mean functional similarity; some of the reported matches may indicate the presence of a conserved domain or motif or simply a common protein structure pattern. Those ESTs identified as fully corresponding to known human genes or proteins are not included in this disclosure. Statistically significant matches are reported in Table 2, together with the length and percent identity or similarity of each alignment.

On the basis of database searches, 609 EST sequences were classified into eight groups as shown in Table 1 (see Example 1 above). Four groups, with 197 or 32% of the sequences, consist of matches to human sequences: repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes. Forty-eight (8%) of the sequences matched non-human entries in GenBank or PIR while 230 (38%) had no significant matches. The remaining 134 (22%) sequences contained no insert or consisted entirely of polyA between the EcoRI cloning sites.

Thirty-six ESTs matched previously sequenced human nuclear genes with more than 97% identity. Four of these ESTs are from genes encoding enzymes involved in maintaining metabolic energy, including ADP/ATP translocase, aldolase C, hexokinase, and phosphoglycerate kinase. Human homologs of genes for the bovine mitochondrial ATP synthase $F_0\beta$-subunit and porcine aconitase were also found (Table 2). Brain-specific cDNAs included synaptophysin, glial fibrillary acidic protein (GFAP), and neurofilament light chain. At least six ESTs are from genes encoding proteins involved in signal transduction: 2',3'-cyclic nucleotide 3'-phosphodiesterase (2 ESTs), calmodulin, c-erbA-α-2, $G_s\alpha$, and Na^+/K^+ ATPase α-subunit. Other ESTs were matches to genes for ubiquitous structural proteins -- actins, tubulins, and fodrin (non-erythroid spectrin). ESTs also document the presence in the hippocampus cDNA library of the ret proto-oncogene, the ras-related gene

rhoB, and one of the chromosome 22 breakpoint cluster region transcripts. Eight ESTs are from genes known to be associated with genetic disorders (Online Mendelian Inheritance in Man). More than half of the human-matched ESTs from Example 1 have been mapped to chromosomes, indicating the bias of GenBank entries toward well-studied genes and proteins.

ESTs without significant GenBank matches were also compared to the ProSite database of recognized protein motifs. Not counting post-translational-modification signatures, fifty-four sequences contained motifs from the database. Some patterns, particularly the "leucine zipper", are found in scores or hundreds of proteins that do not share the functional property implied by the presence of the motif.

Similarities to sequences from other organisms were also detected in the BLAST searches of GenBank and PIR (Table 2). Several ESTs displayed similarity to "housekeeping" genes, including the ribosomal proteins S10 and L30 (rat) and the above glycolytic enzymes. EST00257 (SEQ ID NO:77) shows strong nucleotide sequence similarity to the squid (67%) and Drosophila (70.4%) kinesin heavy chain. Kinesin was first described as a microtubule-associated motor protein involved in organelle transport in the squid giant axon (Vale et al, **Cell 42**: 39 (1985)). Six oncogene-related sequences were also among the cDNA clones sequenced. EST00299 (SEQ ID NO:180) and EST00283 (SEQ ID NO:271) show similarity to several ras-related genes and EST00248 (SEQ ID NO:102) matched the 3' untranslated region of the bovine substrate of botulinum toxin ADP-ribosyltransferase. Similarities with an S. cerevisiae RNA polymerase subunit and Torpedo electromotor neuron-associated protein were also observed. Two ESTs may represent new members of known human gene families: EST00270 matched the three β-tubulin genes with 88-91% identity and EST00271 (SEQ ID NO:248) matched α-actinin with 85% identity at the nucleotide level.

Among the most interesting of the primary sequence relationships was the similarity of ESTs to the Drosophila

genes Notch and Enhancer of split. Nucleotide and peptide alignments of EST00256 (SEQ ID NO:188) and EST00259 (SEQ ID NO:227) with the Drosophila genes have been demonstrated. Both genes are part of a signal cascade encoded by the "neurogenic" genes that are involved in the differentiation of neuronal and epidermal cell lineages in the neuroectoderm of the developing Drosophila embryo (Campos-Ortega, **Trends in Neuro. Sci. 11:** 400 (1988)). It has been proposed that the Enhancer of split protein interacts with a membrane protein that is the product of the Notch gene to convert a developmental signal into an altered pattern of gene expression (id. **J. Mol. Biol. 215:** 403 (1990)). EST00256 (SEQ ID NO:188) matches near the 5' end of the Enhancer of split coding sequence, away from the mammalian G protein ß subunit- and yeast cdc4-like elements (Hartley et al, **Cell 55:** 785 (1988); Klambt et al. **EMBO J. 8:** 203 (1989)). Part of the EST00259 (SEQ ID NO:227) match to Notch in the cdc10/SW16 region that is similar to three cell-cycle control genes in yeast and is tightly conserved in the Xenopus Notch homolog, Xotch. In Drosophila, Enhancer of split is absolutely required for formation of epidermal tissue. Notch contains several epidermal growth factor-like repeats and appears to play a general role in cell-cell communication during development (Banerjee and Zipursky, **Neuron 4:**177 (1990)).

Seven genes were represented by more than one EST. Comparisons of all the ESTs against one another revealed two overlaps of unknown ESTs: EST00233 (SEQ ID NO:32) and EST00234 (SEQ ID NO:8) match in opposite orientations and EST00235 (SEQ ID NO:204) and EST00236 (SEQ ID NO:148) match in the same orientation beginning at the same nucleotide. Five human genes were represented by more than one EST: ß-actin (3), λ-actin (2), α-tubulin (2), α-2-macroglobulin (2), and 2'3'-cyclic-nucleotide-3'-phosphodiesterase (2). Those few instances where two or more ESTs represent different portions of a single cDNA can be readily ascertained when the sequence of the full cDNA insert is determined in accordance with

Example 11.

Table 2: ESTs Identified by Database Matches

SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID
208	EST00250	60K filarial antigen	A28209	PIR	108	56.9
97	EST00289	Aconitase	A35544	PIR	105	90.6
251	EST00370	Actin, other	S10021	PIR	44	51.1
248	EST00271	Actinin, alpha	HUMACTAR	GB	271	85.3
132	EST00110	Agrin	RATAGR	GB	269	82.2
13	EST00255	Cadherins	CADN$HUMAN	SP	41	45.2
188	EST00256	Enhancer of split	A30047	PIR	86	58.6
310	EST00377	Fo ATPase beta subunit, mitochondrial	BOVMTASB	GB	293	85.4
77	EST00257	Kinesin	A35075	PIR	57	86.2
78	EST00258	Kinesin	A35075	PIR	62	47.6
313	EST00276	Lysosomal membrane glycoprotein 1 (LAMP-1)	A31959	PIR	53	46.3
161	EST00247	MARCKS (myristoylated alanine-rich protein kinase	BOVMARCKS	GB	139	83.6
43	EST00371	Maternal G10 protein	S05955	PIR	38	92.3
223	EST00368	Microtubule-associated protein 1B	A33645	PIR	30	54.8
227	EST00259	Notch/Xotch	A35844	PIR	74	85.3
93	EST00287	Processing enhancing protein	S03968	PIR	96	58.8
9	EST00376	Prolyl endopeptidase	PIGPREP	GB	223	83.9
202	EST00298	Protein-tyrosine phosphatase LRP	LRP$MOUSE	SP	62	44.4
38	EST00374	RNA polymerase II 6th subunit (RPO26)	A36352	PIR	72	75.3
37	EST00038	ras p21-like small GTP-binding protein (smg GDS)	BOVSMGGDS	GB	131	89.4
180	EST00299	ras-related proteins	S10493	PIR	51	46.1
102	EST00248	rho H12/ ARH12	BOVBGBRH	GB	195	79.6
301	EST00300	Ribosomal protein L30	R6RT30	PIR	57	96.5
22	EST00301	Ribosomal protein S10	R3RT10	PIR	66	97.0
299	EST00249	smg p25A GDP dissociation inhibitor	A35652	PIR	97	77.5
300	EST00232	Transforming protein (dbl)	TVHUDB	PIR	25	65.4
189	EST00282	trkB	A35104	PIR	33	67.6
187	EST00152	Wilm's tumor-related protein	HUMOM	GB	228	99.6
249	EST00275	Zinc Finger Proteins	S06551	PIR	25	57.7

There is little redundancy in EST sequencing according to the present invention. Of the nuclear-encoded messenger RNAs, the most common ESTs were to the β-actin (approximately 0.6% of the EST clones) and myelin basic protein genes (MBP, approximately 0.5% of the clones). MBP, a highly expressed structural component of nerve tissue (Kamholtz, J., de Ferra, F., Puckett, C., & Lazzarini, R. **Proc. Natl. Acad. Sci., USA 83**: 4962-4966 (1986)), displays four alternate splicing forms, of which it is believed at least two are present among the ESTs reported here. Other common ESTs were Gs-alpha gamma-actin and both a- and alpha-tubulin.

By matching ESTs to known database sequences, a phenotypic characterization of the tissue begins to emerge. Protein superfamilies matched by ESTs were grouped into three broad functional categories to assess the biological spectrum represented by these randomly selected cDNA clones. Structural and metabolic classes comprised about 30% of the ESTs with database matches. Twenty-five percent were involved in regulatory pathways and the remainder were not classifiable. In addition, it is believed that several genes not previously known to be expressed in the brain were matched, including spermine/spermidine acetyltransferase (Casero, R., Celano, P, Ervin, S., Applegren, N., Wiest, L. & Pegg, A. **J. Biol. Chem. 266**: 810-814 (1991)) and osteopontin (Young, M., Kerr, J., Termine, J., Wewer, U., Wang, M., McBride, W. & Fisher, L. **Genomics** 7:491-502 (1990)).

EXAMPLE 3

Mapping of ESTs to Human Chromosomes

Randomly selected ESTs corresponding to Sequence Identification numbers were assigned to chromosomes via PCR (see Table 3). Oligonucleotide primer pairs were designed from EST sequences to minimize the chance of amplifying

through an intron. The oligonucleotides were 18-23 bp in length and designed for PCR amplification using the computer program INTRON (National Institutes of Mental Health, Bethesda, MD). The program is based on the assumptions that: 1) introns are genomic sequences that interrupt the coding and noncoding sequences of genes (Smith, **J. Mol. Evol. 27**:45-55 (1988)); 2) there are consensus sequences for splice junctions (Shapiro, et al., **Nucl. Acids Res. 15**:7155-7174 (1987)); and 3) that 90% of the human genes studied have 3' untranslated regions of mRNA not interrupted by introns in the genomic DNA (Hawkins, **Nucl. Acids Res. 16**:9893-9908 (1988)).

The program evaluates the likelihood that a given GG or CC dinucleotide represents a former exon-intron boundary. Specifically, every input strand is processed by the INTRON program twice, first evaluating the sense mRNA strand, and then processing the complementary or anti-sense strand. The program evaluates each sequence by finding all GG or CC pairs (possible former splice sites), searching for STOP codons in all three reading frames, and analyzing the GG or CC pairs surrounded by stop codons. All regions of the EST that are unlikely to contain splice junctions based on CC content, GG content, and stop codon frequency are then marked by the program in uppercase.

The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich, H.A., **PCR Technology; Principles and Applications for DNA Amplification**, 1992; W.H. Freeman and Co., New York. ESTs were examined for the presence of stop codons in each reading frame and for consensus splice junctions. The presence of stop codons and absence of splice junction sequences are more characteristic of 3' untranslated sequences than of introns. The untranslated sequences are unique to a given gene; thus, primers from these regions are less likely to prime other members of a gene family or pseudogenes.

The primers were used in polymerase chain reactions

(PCR) to amplify templates from total human genomic DNA. PCR conditions were as follows: 60 ng of genomic DNA was used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 uCu of a ^{32}P-labeled deoxycytidine triphosphate. The PCR was performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products were analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the size of the resulting product was equivalent to the EST from which the primers are derived, then the PCR reaction was repeated with DNA templates from two panels of human-rodent somatic cell hybrids; BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR was used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given EST. DNA was isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from EST sequences selected above. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the EST will yield an amplified fragment. ESTs were assigned to a chromosome by analysis of the segregation pattern of PCR products from hybrid DNA templates. For a review of techniques and analysis of results from somatic cell gene mapping experiments. (See Ledbetter et al., **Genomics** 6:475-481 (1990).) The single human chromosome present in all cell hybrids that give rise to an amplified fragment represents the chromosome containing that EST.

The assignment of 81 ESTs and corresponding genes to chromosomes by PCR is shown in Table 3.

Table 3: Assignment of ESTs to Chromosomes by PCR

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
5	EST00012	1	TCCAGGCAATCCCAGAATAG	CTAATTGAGCTCACTGGCCC
57	EST00058	1	CTGTTTGCAAGTTTCAAAGC	GCCATTTCTAACAACCAGAG
64	EST00066	1	GCCATTGTGCTGAATAGAGT	GTTAGTGTTTCCTTAGCAAG
83	EST00079	1	CAGCTAATTGACCTGGGCTA	CAACATGCTCTGAGCTTTAG
83	EST00079	1	GGCAGAGCATAATGAGTATA	CATATGCATATGGTCCCTAT
91	EST00086	1	AGTTTAGATGGAGGGCTGTC	TCTGCCCTAATGCGCAGGCT
105	EST00365	1	CTTAATCACCTCCCTTTTGT	CCTTAGTTGGAGATAAGGTC
109	EST00095	1	AGTCTAATCCTGTACACTTG	CGGGCTTTCTCTGAATTGGT
116	EST00100	1	TTAGAAGTGCCCATGGGAGG	TTTTAAGGCTCTGGAGTGTT
141	EST00118	1	CTCAGAGAAACTTAGGTGAA	CTACAGAATCATTTCACCAG
220	EST00372	1	AAGTTGCACATTGCCCAAGG	ATAGTACTGCAAGGTTATTC
237	EST00187	1	TTACAAATTTCTCTTGACGC	CTGAAGGAGCACAGTTTCTC
242	EST00192	1	GGATCAGATAATCAAACAGG	GCTTAGGATATGAATGCATA
259	EST00202	1	GCATCACAGTTTAACTGAGG	CTACATATTTGTGCCTCCTT
269	EST00293	1	CTGTTGCTGTGCAGTAGCTT	CTTTTGACCCAGTGAAACTT
299	EST00249	1	GATCATGCAGACGTAGATAT	CCAACTCCTGCCAGATCATT
16	EST00021	2	CAGGCAAGTTTCTTCCAGGA	TCAGACCCATGGTCAGCTT
8	EST00234	2	TAGAAGGCAAACTATGTCCC	GGTTGAGGATTGGCTTTTAC
36	EST00037	2	AGCCAGAAGGCTGCTTAAAG	GCAGTGAACCAGTACTCCTA
123	EST00106	2	GTCTAATTTGTAACCTTCAG	GATAGATTGTATAAGAAGCC
192	EST00155	2	GATTTATGTCTGGGAACTAA	GCAGCATGTGAAAGAATGAT
200	EST00162	2	TTTAATGGGTGGTGGGAGCT	CGATGCACATCCTTCTCCAT
284	EST00216	2	CCTAAGAATTCGTTTGGCTC	GTCTGGCACATAATAGATTTG
102	EST00248	3	ATACTACATCTAGTCTGG	TTACAGTTCTGTGGTTTC
167	EST00138	3	AAACAGCTGCGGAGTACA	AAAGGATCCTCCACTCCAGA
12	EST00274	3	CCTAGCAAACTCATACACAC	CATAAGTGAATGGACACAGG
60	EST00062	3	ACACATTAACGGTGCTGCAG	GGAATCAGCCCTTGAGGACT
77	EST00257	3	AAGCTCACAACGCAGATCTG	CTGGAACAGCTTACAAAGGT
107	EST00093	3	ATTGAACTCTGTCAACAGTG	TGTAAAACAAAGGCCAAACT
108	EST00094	3	AL2-GCAGGATGTCAGTCTTTTGAG	AGCACACATTATCTACCACGGC
37	EST00038	4	AACTTCGCAGTCATGAGAAC	TGTATCGGGCAGTTCTCAG
6	EST00013	4	CACATGTTCTCCCTCTTTCA	GCATTTTGGAGCTCTTCCGT
37	EST00038	4	AL2-GGAAGTACAGGATTTGGC	TTAGAGATGGGATGATGCCG
31	EST00033	5	TGGGTACCCTAAGGTGTTTG	GACTAATCTAAGGTCTAGG
28	EST00030	5	AGATAAGTTAGGAAGCTGGT	ACTCACTGCTAGTATCATCC
59	EST00061	5	AAAGTTTCTTAGCACCCCCC	CAGACTTTGACAAAAGAATC
74	EST00073	5	ATCAGACACGTGGCAGGGTT	AAGTCCCTGAGGGTGCAGAA
121	EST00104	5	TGAAGGCAGCTGCTAAATCT	GGATGTATTGATCTGACTCA
149	EST00123	5	ATACTGTCAACGGAGGGTGA	GTCTGCAGGTTTCTCCTTGA
235	EST00185	5	TTACTGTCCCATCAGATATC	TACACTCTTAAGAAGGTATG
23	EST00026	5	CCTGCAGTGACACTTAACAT	CTGCTCACCTGAAATTGATAC
121	EST00104	5	AL2-CAGATCAATACATCCTCTGGG	CTGTGCAGTGGTGAGTAAAAGG
1	EST00007	6	TAGTTGATGGTCTGGGTTAT	GAAATCCCAGGGAGACAATG
19	EST00023	6	CAACTTACATTAGGGGTTTG	GACCTCATTAGAAGAGCCCA
155	EST00129	6	GGAAGCTGCCATATAAGCTC	TCAGTGTCGTACAATCTACC
224	EST00356	6	GCTGTATGTTAACCCTTTGT	TGGAACCCTCAAACACTGCT
288	EST00219	6	ACTTTCATGTTGAGAAGTAT	ATCTAGCTGAAACATTGCTG
22	EST00301	6	CTCCGTGATTACCTTCATCT	TTGTAGGTATCTCTGTCAGCT
207	EST00167	7	GGTGCTACTTTGTGAATGCT	AGCAATGTGATTTTGTAGG
137	EST00272	7	AGTGGTCACTATCTACATGG	GATTCAGAATTACTAAGCCG
292	EST00223	8	TGCAGCAGTGACCATGAGAA	ATCATCTTTCCACGCGGCTT

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
134	EST00375	9	TCTGGGCTTCTGTGGTTCAA	CTGGCTGCTCAGCAACTCAT
20	EST00024	10	AGCTGTTCCTGAGAGATGCA	CCTTGTGAAGAAAGACTTTC
157	EST00131	10	TCAGCAACAGGTCACTTTGG	CTAAGCATCTGCATGTCCAG
172	EST00142	10	TACTAGCATTTCTTACTCTC	TATGCTGATTGTTTGCACTC
250	EST00197	10	GGTGATTAGAGAGTCTGTTG	GAACTCTGTAGTGTTCTAAA
133	EST00111	11	GGAAATTAGGCTTAGCTCAC	GTGCAGAATACTTAGAGTCC
178	EST00294	11	GTTTGAAGGAAGTGATTTCC	TAGGGCCACCTCCAGTTCAT
10	EST00016	11	GTCTTTGGATTCTACGTAGA	CGATAATGACATTTCTTCTGG
126	EST00109	11	AL2-CTAACCACAACCCACACATTG	CCTCAGCACAAGAGAAGAATGG
7	EST00014	12	AACTTGCAACATAAATACTAG	GAGCAATGATTTCTAACAGT
254	EST00200	13	TTGTGTACTGTCTGATAGAC	TAAGCCATGGGCATCTATAA
170	EST00295	14	GGTGCTTAAGGCCACTTTTG	CTTAGAGGATCATAGGTCTG
255	EST00201	14	CCAGGAGAGTAAGAAGATCA	GCAGAGTTGAATATGAACCT
290	EST00221	14	GTGCCAAGATGGCTCATGTA	GTATAGCTTTAAGCCAGTTC
293	EST00224	14	AATGCATTATGCCTGGTCTT	GGAAAAGTCTAGAACTTAGT
315	EST00008	14	AAGCTGGCTGGGAAATGTTC	GTCATGCTAGTAAACTTACAC
95	EST00088	15	GTGACAGACCATGTCTATTG	AAGTGAGCGATTGCACCTTC
205	EST00165	15	AGGATGACCTGAGTGAGCTG	CCATGGCAGCAAGGAACTCT
33	EST00034	16	TGTGTGAAAGGGAGTCTTGT	CCATTTTGACTGTTCCATAG
247	EST00279	16	TGGCTAGGGCAGGCCTTAAA	GAGAAGAATATCAAATGGGG
18	EST00373	16	CCATCTGTGTCCCAATTAAGC	AGGGAAGAAGTCTAGAGCGA
68	EST00068	17	CAAAGACGGGAGACGAATGA	ACTGGAACGCGTGGCCTATG
84	EST00080	19	AGAGATGTCAGTCCATTATC	CTATTCCACCTTACTCAAGG
223	EST00368	19	CATCATGTCGGAGACGCATT	TGGATGACCTGAGTCTGCAG
21	EST00025	20	AGTTCTGGAGGCTAGGAGTT	ATGTAAGGACCCCTAGATGG
210	EST00168	20	TGTCAACTTCCCTTTGGCCT	GAAGCTTGCTCATTCAGGAA
136	EST00113	20	AL2-TCGGAGAAGTTGCAGTTTCTG	GTTAAAAGCTGTTAGACGGGGC
120	EST00103	21	CACTGACTGACTCCTCTTTA	GGAACCGTAACTCTCCATAG
313	EST00276	X	ATTGACCTTCAATGTAATAA	TTGGATTGGGCAAAATAG
162	EST00133	X	ATGTGAGCATCTATACCTGC	AATGAAGGCATGAGAATAGG

Abbreviations: AL2: Amino-Link-2 Fluorescent Tag, Chr.: Chromosome.

The foregoing techniques have been used to further localize 6 ESTs and their associated genes to precise locations onto chromosome 6 or chromosome X, as reflected in Table 4 (in Example 5 below), using sublocalization techniques that employ somatic cell hybrids. ESTs were used as hybridization probes and mapped to other chromosomes using techniques disclosed in Example 5. Somatic cell hybrids were prepared that contained defined subsets of chromosomes 6 and X. Methods for preparing and selecting somatic cell hybrids are known in the art. For a review of an exemplary procedure to generate somatic cell hybrids containing the short arm of human chromosome 6, see Zoghbi, et al., **Genomics 9(4)**:713-720 (1991). For a general review of somatic cell hybridization see Ledbetter et al.(supra). The hybrids were processed to obtain DNA and analyzed by PCR and by fluorescence in situ hybridization. SEQ ID NOs 19, 22, 1, 224, 288 mapped to chromosome 6, while SEQ ID NO 162 mapped to chromosome X using somatic cell hybrids.

EXAMPLE 4

Mapping of All ESTs to Human Chromosomes

The procedure of Example 3 is repeated for all of the ESTs from Example 1 not previously mapped to human chromosomes. Data are generated corresponding to the data in Table 3 for all of the unmapped ESTs. As previously mentioned, virtually all of the ESTs will map to a unique chromosomal location. The inability of any ESTs to localize to a unique location will be readily ascertainable during the mapping process.

EXAMPLE 5

Alternative Technique for Mapping to Chromosomes Mapping of ESTs to chromosomes using fluorescence in situ hybridization

This technique is used to map an EST to a particular location on a given chromosome. Cell cultures, tissue, or whole blood can

be used to obtain chromosomes.

0.5 ml. of whole blood is added to RPMI 1640 and incubated 96 hours in a 5%CO_2/37°C incubator. 0.05 ug/ml colcemide is added to the culture one hour before harvest. Cells are collected and washed in PBS. The suspension is incubated with a hypotonic solution of KCl added dropwise to reach a final volume of 5 ml. The cells are spun down and fixed by resuspending the cells in methanol and glacial acetic acid (3:1). The cell suspension is dropped onto glass slides and dried.

The slides are then treated with RNase A and washed then dehydrated in a series of increasing concentrations of ethanol.

The EST to be localized is nick-translated using fluorescently labeled nuclectide (Korenberg, Jr., et al., **Cell 53(3)**:391-400 (1988)). Following nick translation, unincorporated label is removed by spin dialysis through Sepharose. The probe is further extracted with phenol-chloroform to remove additional protein. The chromosomes are denatured in formamide using techniques known in the art and the denatured probe added to the slides. Following hybridization, the cells are washed. The slides are studied under a fluorescent microscope. In addition, the chromosomes can be stained for G-banding or Q-banding using techniques known in the art.

The resulting metaphase chromosomes have fluorescent tags localized to those regions of the chromosome that are homologous to the EST. Thus, a particular EST is localized to a particular region on a given chromosome. For a review of the technique, see Verma et al., **Human Chromosomes: A Manual of Basic Techniques.** Pergamon Press, NY (1988), which is hereby incorporated by reference.

Table 4: Precise Chromosomal Localization of ESTs

SEQ ID	EST#	Map Location
19	EST00023	6p
22	EST00301	6p
1	EST00007	6q
224	EST00356	6q
288	EST00219	6q
162	EST00133	Xp11.21 - Xp21.2

EXAMPLE 6

Automated DNA Sequencing Accuracy

ESTs that match human sequences in GenBank are excellent tools for the analysis of the accuracy of double-strand automated DNA sequencing. EST/GenBank matches from a number of clones were examined for the number of nucleotide mismatches and gaps required to achieve optimal alignment by the Genetics Computer Group (GCG) program BESTFIT (Devereux et al, **Nucleic Acids Research 12**: 387 (1984)). The number of mismatches, insertions and deletions was counted for each hundred bases of the sequence (Table 5). As expected, the sequence quality was best closest to the primer and decreased rapidly after about 400 bases. The number of deletions and insertions relative to the GenBank reference sequence increased five- to ten-fold beyond 400 bases, while the number of mismatches doubled. The average accuracy rate for individual double-stranded sequencing runs was 97.7% to 400 bases.

TABLE 5. Accuracy Of Single-Run Double-Stranded Automated Sequencing

Bases from Primer	Mismatches/ Ambiguities[1]	Gaps Insertions[1]	Percent Deletions[1]	Aligned Accurate	Bases
101 - 200	1.45	0.18	0.19	98.2	8,800
201 - 300	1.72	0.25	0.11	97.9	8,130
301 - 400	2.07	0.98	0.37	96.6	5,404
>400	3.53	2.63	1.06	92.8	3,197

ESTs statistically identical to known human sequences and those matching mitochondrial and ribosomal genes were aligned with sequenced from GenBank using the GCG program BESTFIT. The first 85 nucleotides was polylinker sequence which was not aligned with the pBluescript SK reference sequence. Tabulation of errors began 15 bases into the BESTFIT alignment and thus is reported beginning with bases 101-200. [1]Error rates are reported as number of mismatches, insertions, or deletions per hundred aligned bases. "Mismatches" Includes ambiguous base calls.

EXAMPLE 7

Probability of ESTs Containing Coding Sequences

The ESTs of the present invention were statistically evaluated using the coding-region prediction program CRM via the GRAIL server (Uberbacher, E. & Mural, R. **Proc. Natl. Acad. Sci. USA, 88:** 11261-5 (1991)). The CRM program uses a neural network to combine results from several different coding regions by looking at different 6 bp sequences found in coding exons and in introns. The program additionally conducts reading frame searches and assesses randomness at the third position of codons. This protocol categorizes sequences as having an excellent, good, marginal, or poor probability of containing coding regions. The results are reported in Tables 6-9. There were 32 ESTs categorized as "excellent" (Table 6); 14 categorized as "good" (Table 7); 13 categorized as "marginal" (Table 8); and 213 categorized as "poor" (Table 9). These results indicate that most ESTs of the present invention comprise noncoding regions.

Table 6: ESTs with Excellent Probability of Containing Coding Sequence

SEQ ID#	EST#
1	EST00014
15	EST00020
48	EST00291
62	EST00084
66	EST00087
75	EST00074
95	EST00280
106	EST00092
108	EST00094
114	EST00098
115	EST00099
124	EST00107
128	EST00252
156	EST00130
164	EST00135
166	EST00137
174	EST00296
179	EST00145
183	EST00146
201	EST00163
205	EST00165
215	EST00172
230	EST00181
253	EST00189
263	EST00203
285	EST00368
270	EST00207
271	EST00283
273	EST00208
278	EST00211
281	EST00214
289	EST00286

Table 7: ESTs with Good Probability of Containing Coding Sequence

SEQ ID=	EST=
20	EST00024
72	EST00071
82	EST00078
88	EST00084
137	EST00272
177	EST00328
193	EST00156
200	EST00162
218	EST00175
228	EST00179
247	EST00279
264	EST00204
267	EST00297
296	EST00228

Table 8: ESTs with Marginal Probability of Containing Coding Sequence

SEQ ID#	EST#
11	EST00018
12	EST00274
24	EST00027
45	EST00364
79	EST00076
90	EST00302
110	EST00096
144	EST00120
145	EST00121
192	EST00155
222	EST00177
234	EST00184
277	EST00212

Table 9: ESTs with Poor Coding Probability

SEQ ID#	EST#	SEQ ID#	EST#	SEQ ID#	EST#	SEQ ID#	EST#
1	EST00007	100	EST00090	195	EST00158	295	EST00226
2	EST00009	101	EST00091	196	EST00159	297	EST00230
3	EST00010	103	EST00317	197	EST00160	298	EST00231
4	EST00011	104	EST00354	198	EST00161	302	EST00303
5	EST00012	105	EST00365	199	EST00277	303	EST00348
6	EST00013	107	EST00093	203	EST00164	304	EST00307
8	EST00234	109	EST00095	204	EST00235	305	EST00308
10	EST00016	111	EST00281	206	EST00166	306	EST00309
14	EST00019	112	EST00318	207	EST00167	307	EST00312
16	EST00021	113	EST00097	209	EST00331	308	EST00314
17	EST00022	116	EST00100	210	EST00168	309	EST00174
18	EST00373	117	EST00319	211	EST00332	315	EST00008
19	EST00023	118	EST00101	212	EST00169		
21	EST00025	119	EST00102	213	EST00170		
23	EST00026	120	EST00103	214	EST00171		
25	EST00028	121	EST00104	216	EST00173		
27	EST00029	122	EST00105	219	EST00176		
28	EST00030	123	EST00106	220	EST00372		
29	EST00031	125	EST00108	221	EST00359		
30	EST00032	126	EST00109	224	EST00356		
31	EST00033	127	EST00320	225	EST00178		
32	EST00233	129	EST00321	226	EST00333		
33	EST00034	130	EST00355	229	EST00180		
34	EST00035	131	EST00322	231	EST00334		
35	EST00036	133	EST00111	232	EST00182		
36	EST00037	134	EST00375	233	EST00183		
39	EST00039	135	EST00112	235	EST00185		
40	EST00040	136	EST00113	236	EST00186		
41	EST00041	138	EST00114	237	EST00187		
42	EST00042	139	EST00116	238	EST00188		
46	EST00044	140	EST00117	239	EST00189		
47	EST00046	141	EST00118	240	EST00335		
49	EST00047	142	EST00323	241	EST00191		
50	EST00048	143	EST00119	242	EST00192		
51	EST00049	146	EST00122	243	EST00193		
52	EST00052	147	EST00292	244	EST00194		
53	EST00054	148	EST00236	245	EST00347		
54	EST00055	149	EST00123	246	EST00196		
55	EST00056	150	EST00124	250	EST00197		
56	EST00057	151	EST00125	252	EST00198		
57	EST00058	152	EST00126	254	EST00200		
58	EST00059	153	EST00127	255	EST00201		
59	EST00061	154	EST00128	256	EST00345		
60	EST00062	155	EST00129	257	EST00337		
63	EST00065	157	EST00131	259	EST00202		
64	EST00066	158	EST00132	260	EST00357		
67	EST00351	159	EST00325	261	EST00338		
68	EST00068	160	EST00326	262	EST00339		
69	EST00360	162	EST00133	265	EST00205		
71	EST00070	163	EST00134	266	EST00206		
73	EST00072	165	EST00136	272	EST00340		
74	EST00073	167	EST00138	274	EST00268		
76	EST00075	168	EST00140	275	EST00209		
80	EST00077	169	EST00141	278	EST00342		
81	EST00315	170	EST00295	279	EST00213		
83	EST00079	171	EST00327	280	EST00343		
84	EST00080	172	EST00142	283	EST00215		
85	EST00081	173	EST00143	284	EST00216		
86	EST00082	175	EST00144	286	EST00217		
87	EST00083	178	EST00294	287	EST00218		
89	EST00085	182	EST00329	288	EST00219		
91	EST00086	184	EST00149	289	EST00220		
92	EST00087	185	EST00150	290	EST00221		
94	EST00353	186	EST00151	291	EST00222		
95	EST00088	190	EST00153	292	EST00223		
96	EST00089	191	EST00154	293	EST00224		
99	EST00316	194	EST00157	294	EST00225		

EXAMPLE 8

Functional Groupings of ESTs and Corresponding Genes

By matching new human ESTs to known sequences from other species, the apparent function of the gene corresponding to the EST can be ascertained. The data generated in Example 2 have been used to categorize 28 of the ESTs of the present invention, and their corresponding genes, into predicted functional groups. (These 28 are ESTs with database matches to sequences from other species for which a function was known.) Two different grouping schemes have been used.

The first scheme separates the sequences into three broad categories: metabolic; regulatory; and structural. These groupings are set out in Table 10.

The second grouping scheme separates the sequences into 13 specific categories: cell surface proteins; developmental control; energy metabolism; kinases and phosphatases; oncogenes; other metabolism-related polypeptides; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transporters; transcription, translation, and subcellular localization; and transcription factors. These groupings are set out in Table 11.

Table 10: Three-Class Functional Groupings of ESTs

SEQ ID	EST#	Group	Putative Identification
97	EST00289	M	Aconitase
310	EST00377	M	Fo ATPase beta subunit, mitochondrial
93	EST00287	M	Processing enhancing protein
9	EST00376	M	Prolyl endopeptidase
38	EST00374	M	RNA polymerase II 6th subunit (RPO26)
301	EST00300	M	Ribosomal protein L30
22	EST00301	M	Ribosomal protein S10
188	EST00256	R	Enhancer of split
161	EST00247	R	MARCKS (myristoylated alanine-rich protein kinase
227	EST00259	R	Notch/Xotch
202	EST00298	R	Protein-tyrosine phosphatase LRP
300	EST00232	R	Transforming protein (dbl)
37	EST00038	R	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	R	rho H12/ ARH12
299	EST00249	R	smg p25A GDP dissociation inhibitor
189	EST00282	R	trkB
43	EST00371	R	Maternal G10 protein
187	EST00152	R	Wilm's tumor-related protein
249	EST00275	R	Zinc Finger Proteins
208	EST00250	S	60K filarial antigen
251	EST00370	S	Actin, other
248	EST00271	S	Actinin, alpha
132	EST00110	S	Agrin
77	EST00257	S	Kinesin
78	EST00258	S	Kinesin
313	EST00276	S	Lysosomal membrane glycoprotein 1 (LAMP-1)
223	EST00368	S	Microtubule-associated protein 1B
311	EST00270	S	Tubulin, beta

Group Key: M: Metabolic, R: Regulatory, S: Structural

Table 11: Thirteen-Class Functional Groupings of ESTs

SEQ ID	EST#	Group	Putative Identification
208	EST00250	CS	60K filarial antigen
313	EST00276	CS	Lysosomal membrane glycoprotein 1 (LAMP-1)
188	EST00256	DC	Enhancer of split
43	EST00371	DC	Maternal G10 protein
227	EST00259	DC	Notch/Xotch
97	EST00289	EM	Aconitase
310	EST00377	EM	Fo ATPase beta subunit, mitochondrial
202	EST00298	KP	Protein-tyrosine phosphatase LRP
300	EST00232	OG	Transforming protein (dbl)
37	EST00038	OG	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	OG	rho H12/ ARH12
9	EST00376	PI	Prolyl endopeptidase
189	EST00282	RT	trkB
251	EST00370	SC	Actin, other
132	EST00110	SC	Agrin
77	EST00257	SC	Kinesin
78	EST00258	SC	Kinesin
223	EST00368	SC	Microtubule-associated protein 1B
311	EST00270	SC	Tubulin, beta
161	EST00247	ST	MARCKS (myristoylated alanine-rich protein kinase
299	EST00249	ST	smg p25A GDP dissociation inhibitor
93	EST00287	TT	Processing enhancing protein
38	EST00374	TT	RNA polymerase II 6th subunit (RPO26)
301	EST00300	TT	Ribosomal protein L30
22	EST00301	TT	Ribosomal protein S10
187	EST00152	TX	Wilm's tumor-related protein
249	EST00275	TX	Zinc Finger Proteins

Group Key: CS: Cell Surface, DC: Developmental Control, EM: Energy Metabolism, KP: Kinases and Phosphatases, OG: Oncogenes, PI: Peptidases and Peptidase Inhibitors, RT: Receptors, SC: Structural and Cytoskeletal, ST: Signal Transduction, TT: Transcription, Translation, and Subcellular Localization, TX: Transcription Factors.

EXAMPLE 9

cDNA Libraries Generated From Specific Genomic DNA by Exon Expression & Amplification

Exon amplification is used to express potential exons from genomic DNA in a recombinant vector that contains some of the signals necessary for splicing. If an exon is present in the proper orientation in the vector, that exon will be spliced in a mammalian cell and will become part of the mRNA of that cell. The exon splice-product can be purified from other mRNA in the cell by conversion of the mRNA to cDNA and selective amplification of the recombinant splice-product cDNAs. Cosmid DNA from human chromosome 19q13.3 is digested with BamHI or BamHI/BglII restriction enzymes. The fragments generated are collected and size specifically cloned into an expression vector (Buckler, et al. **Proc. Nat'l. Acad. Sci. USA,** 88:4005-4009 (1991)). After transfection by electroporation of these constructs into COS cells, RNA transcripts are generated using the SV40 early promoter and a polyadenylation signal derived from SV40 both present in the expression vector. When a fragment of genomic DNA contains an entire exon with flanking intron sequence in the sense orientation, the exon should be retained in the mature poly(A)+ cytoplasmic RNA. Therefore, the mRNA is used as template for cDNA synthesis using reverse transcriptase and vector-priming. Subsequently, the cDNAs are amplified by vector-priming using PCR. A fraction of this first PCR product is reamplified using internal vector-primers containing terminal cloning sites. These products are end-repaired with T4 DNA polymerase, digested with the appropriate restriction enzymes, gel purified and cloned into pBluescript vectors. The constructs are transfected into XL1-Blue competent cells and plated on LB/X-gal/IPTG/ampicillin plates. When multiple cosmids or YAC clones are used as the source DNA, a pool of specific expressed exons is obtained as a cDNA

library.

EXAMPLE 10

PCR Amplification from Predicted Exons

Computational analyses can be applied to genomic DNA sequences to predict protein coding regions. The coding region prediction program CRM (E. Uberbacher and R. Mural, **Proc. Natl. Acad. Sci. USA 88**:11261-5 (1991)) finds open reading frames and classifies them according to their probability of being coding regions. These regions are subsequently examined using the GM program (C. Fields and C. Soderlund, **Comp. Applic. Biosci. 6**: 263, 1990), which predicts intron-exon structure. PCR primers are then designed to amplify the predicted exons and used to test human cDNA libraries (for example, fetal brain or placental libraries) for the presence of these putative exons using a PCR assay.

This strategy has been successfully applied in two large scale genomic sequencing projects, the Huntington's locus of human chromosome 4p16.3 (McCombie, et al., submitted) and human chromosome locus 19q13.3 (Martin-Gallardo, et al., submitted).

EXAMPLE 11

Complete Sequence of EST Clone Inserts

There are a number of methods known to those with skill in the art of molecular biology, to obtain sequence information from the cDNAs corresponding to the EST sequences. Procedures for these methods are provided in Basic Methods in Molecular Biology (David et al. supra). One way to acquire more information about the cDNA from which an EST was derived is to sequence the remainder of the cDNA clone. The complete sequence of the inserts of four EST clones (representing SEQ ID NOs 188, 189, 203, and 207) was determined using Exonuclease III deletions. Briefly, EST clones were digested with the restriction enzymes SalI and NpnI or PstI and BamHI

(for deletions from the Forward primer and Reverse primer ends of the insert, respectively). The KpnI and PstI enzymes leave 3' sticky ends following digestion, which Exonuclease III is unable to bind. This results in unidirectional deletions into the cDNA insert leaving the vector sequence undisturbed. After addition of Exonuclease III to the Forward and Reverse deletion reactions, aliquots of the reaction were removed at defined time intervals and the reaction was stopped to prevent further deletion. S1 nuclease and Klenow DNA polymerase were added to create blunt ended fragments suitable for ligation.

Samples for each time point was purified by electrophoresis through an agarose gel and religated. Two to four representative clones from each time point in each direction were sequenced to give between 200 and 400 base pairs of sequence data. Careful selection of deletion conditions and time points allow a deletion series of approximately 100-200 base pairs difference in length at each consecutive time point. Sequence fragments were reassembled into a redundant contiguous sequence using the INHERIT software from Applied Biosystems, Inc. (Foster City, CA). In this way, the complete insert from these four cDNA clones was sequenced on both strands to an average redundancy between three and four (each base was sequenced between three and four times, on average).

EXAMPLE 12

Determining Reading Frame, Orientation, Coding Regions: ESTs and Complete cDNA Sequences

Once the complete cDNA sequence has been determined in accordance with Example 11, the reading frame, orientation, and coding regions are determined by computer techniques. (The complete coding region is considered to be the largest open reading frame from a methionine to a stop codon.)

Specifically, the CRM program on the GRAIL server is used as explained in Example 7 to determine probable coding regions. This information is supplemented by location of start and stop codons. Where possible, the results of the CRM

analysis are validated by comparison of the cDNA sequence to known sequences using database matching, in accordance with Example 2. If a match of 50% (or even less) is found in any particular reading frame and orientation, this serves to verify corresponding CRM results. Alternatively, database matches can be used to determine reading frame and orientation without use of the CRM program. Of course, if the cDNA is derived from a directional library, the probable orientation is already known.

EXAMPLE 15

Preparation of PCR Primers and Amplification of DNA

The EST sequences and the corresponding cDNA sequences and genomic sequences may be used, in accordance with the present invention, to prepare PCR primers for a variety of applications. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. The procedure of Example 3 is repeated using the desired EST, or using the corresponding cDNA or genomic DNA sequence from Example 11. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. When screening cDNA, introns are of no concern; however, when screening genomic DNA, primers should be selected to avoid reading across introns, which usually are too large to amplify. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

EXAMPLE 14

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers derived from a number of the sequences of Example 1, 9, 10 and/or 11 is then utilized in accordance with Example 10 to obtain DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a suspect. Each of these identification DNAs is then sequenced, and a simple database comparison determines the differences, if any, between the sequences from the suspect and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 15

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of sequences from Examples 1, 9, 10 and/or 11. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 13. Each of these DNA segments is sequenced, using the methods set forth in Example 1. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at

any later time to absolutely correlate tissue or other biological specimen with that individual.

EXAMPLE 16

Southern Blot Forensic Identification

The procedure of Example 15 is repeated to obtain a panel of from 10 to 2000 amplified sequences from an individual and a specimen. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis et al. (**Basic Methods in Molecular Biology**, 1986, Elsevier Press. pp 62-65).

A panel of ESTs or complete cDNA sequences from Examples 1, and/or 11, or fragments thereof of at least 15 bases, are radioactively or colorimetrically labeled using end-labeled oligonuclectides derived from the ESTs, nick translated sequences or the like using methods known in the art and hybridized to the Southern blot using techniques known in the art (Davis et al., supra). Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of ESTs will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of EST probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 17

Dot Blot Identification Procedure

Another technique for identifying individuals using the sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length were synthesized that correspond to sequences from the ESTs. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labelled with ^{32}P using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting about 50 ng cDNA of preferably at least 10 sequences corresponding to a variety of the Sequence ID NOs provided in Table 7 onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the EST clone sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et al. supra). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood et al., **Proc. Natl. Acad. Sci. USA 82(6)**:1585-1588 (1985) which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individuals.

EXAMPLE 18

Alternative "Fingerprint" Identification Technique

EST sequences and the corresponding complete cDNA sequences can be used to create a unique fingerprint for an individual. Thus pools of EST sequences can be used in forensics, paternity suits or the like to differentiate one individual from another.

Entire EST sequences can be used; similarly oligonucleotides can be prepared from EST sequences. In this example, 20-mer oligonucleotides are prepared from 200 EST sequences using commercially available oligonucleotide services such as Oligos Etc., Wilsonville, OR. Patient cell samples are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred using Southern blotting techniques onto nitrocellulose.

10 ng of each of the oligos are pooled and end-labeled with ^{32}P. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the representative number of EST sequences can be varied for additional accuracy or clarity.

EXAMPLE 19

Identification of genes associated with hereditary diseases

This example illustrates an approach useful for the association of EST sequences with particular phenotypic characteristics. In this example, a particular EST is used as a test probe to associate that EST with a particular phenotypic characteristic.

A search of Mendelian Inheritance in Man (supra) revealed 6p21 to be a very gene rich region of the genome containing several known genes and several diseases for which genes have not been identified. Any cDNA encoded by an EST located in this region would thus becomes an immediate candidate for each

of these genetic diseases.

Cells from patients with these diseases are isolated and expanded in culture. PCR primers from the EST sequences are used to screen genomic DNA and RNA or cDNA from the patients. ESTs that are not amplified in the patients can be positively associated with a particular disease by further analysis.

EXAMPLE 20

Identification of a gene associated with Angelman's disease

Angelman's disease (AD) is characterized by deletions on the long arm of chromosome 15 (15q11q13) (Williams et al. **Am. J. Med. Genet.** 32:339-345 (1989) hereby incorporated by reference). The symptoms of the disease include developmental delay, seizures, inappropriate laughter and ataxic movements. These symptoms suggest that the disorder is a neurologic deficiency. This prophetic example illustrates how ESTs, preferably obtained from a cDNA library from human brain, may be used in identifying the defective gene or genes associated with Angelman's Disease. (The example is based on analogous work with genomic DNA, rather than cDNA and ESTs, in identifying the genetic defect associated with Angelman's Disease.) This example also illustrates how EST sequences may generally be used for identifying gene sequences associated with an inherited disease that is mapped to a chromosome location.

ESTs are screened using techniques described in Example 3 and Example 5 to identify those ESTs that localize to the long arm of chromosome 15 and preferably localize to chromosome 15 bands 15q11q13 from normal patients. ESTs that bind to the long arm of chromosome 15 are hybridized to chromosome 15 from AD patients. These studies are preferrably performed using either fluorescence in situ hybridization or using somatic cell hybrids that contain fragments from the long arm of chromosome 15 from AD patients. Those chromosome 15-specific ESTs that do not map to chromosome 15 from AD

patients are useful as markers for Angelman's Disease and can be incorporated into diagnostics for genetic screening. These ESTs are associated with chromosome deletions present in Angelman's disease. Identification of the gene associated with these AD negative ESTs and an analysis of the polypeptides encoded by the genes from normal patients is essential for providing gene or other therapies for AD patients.

Genetic diseases are not always accompanied by gene deletions. Therefore, it is also important to use the ESTs that bind to bands 15q11q13 from AD patients as tools to identify the polymorphisms present within the disease population. Restriction fragment length polymorphism (RFLP) analysis can be performed on patient cells from AD disease or from somatic cell hybrids created using the long arm of chromosome 15. For a review of RFLP techniques see Donis-Keller et al. (**Cell 51**:319-337 (1987) hereby incorporated by reference). DNA is isolated from the somatic cell lines or from cells from AD patients. The DNA is digested with one or more restriction enzymes according to techniques of Donis-Keller et al. The resulting fragments are separated by gel electrophoresis, denatured, transferred to nitrocellulose and hybridized with the selected radio-labeled ESTs that localize to the region of interest. The autoradiographic pattern is compared both to a number of AD patients and to normal patients. Common patterns of EST hybridization in AD patients that are not present in normal patients indicates that the genes associated with these ESTs are candidate genes affected by AD.

cDNA libraries are prepared from the somatic cell hybrids from AD patients. Libraries are prepared using Lambda Zap II Library Kits (Stratagene, La Jolla, California) or other commercially available library kits. The ESTs of interest are used as probes to identify those bacterial colonies carrying genes corresponding to the EST probes. Positive clones are sequenced and the sequences are compared to homologous gene sequences derived from normal patients.

Alterations, including deletions and substitutions, within gene sequences, associated with bands 15q11q13, are thus positively identified and associated with AD disease. Wagstaff et al. were able to identify deletions and substitutions in sequences encoding the $GABA_A$ receptor protein subunit from patients with Angelman's disease (**Am. J. Hum. Genet. 49**:330-337, (1991)). It is likely that other genes will additionally be associated with the disease.

EXAMPLE 21

Preparation and Use of Antisense Oligonucleotides

Antisense RNA molecules are known to be useful for regulating translation within the cell. Antisense RNA molecules can be produced from EST sequences or from the corresponding gene sequences. These antisense molecules can be used as diagnostic probes to determine whether or not a particular gene is expressed in a cell. Similarly, the antisense molecules can be used as a therapeutic to regulate gene expression once the EST is associated with a particular disease (see Example 20).

The antisense molecules are obtained from a nucleotide sequence by reversing the orientation of the coding region with regard to the promoter. Thus, the antisense RNA is complementary to the corresponding mRNA. For a review of antisense design see Green et al., **Ann. Rev. Biochem. 55**:569-597 (1986), which is hereby incorporated by reference. The antisense sequences can contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of the modifications are described by Rossi et al., **Pharmacol. Ther. 50(2)**:245-254, (1991).

Antisense molecules are introduced into cells that express the gene corresponding to the EST of interest in culture. In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that

the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabelling. The antisense molecule is introduced into the cells by diffusion or by transfection procedures known in the art. The molecules are introduced onto cell samples at a number of different concentrations preferably between $1x10^{-10}$M to $1x10^{-6}$M. Once the minimum concentration that can adequately control translation is identified, the optimized dose is translated into a dosage suitable for use in vivo. For example, an inhibiting concentration in culture of $1x10^{-7}$ translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals.

The antisense can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as oligonucleotide contained in an expression vector such as those described in Example 23. The antisense oligonucleotide is preferably introduced into the vertebrate by injection. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate. It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to bind and cleave its target. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al.

EXAMPLE 22

Preparation and use of Triple Helix Probes

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for

studying alterations in cell activity as it is associated with a particular gene. The EST sequences or complete sequences of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with a particular gene. Similarly, a portion of the EST or corresponding gene sequence can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful. However, homopyrimidine sequences can also inhibit gene expression. Thus, both types of sequences from either the EST or from the gene corresponding to the EST are contemplated within the scope of this invention. Homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. As an example, 10-mer to 20-mer homopyrimidine sequences from the ESTs can be used to inhibit expression from homopurine sequences. SEQ ID NOs such as 282 and 240 contain homopyrimidine 15-mers. Moreover the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (**Science 245**:967-971 (1989), which is hereby incorporated by this reference).

The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis. The sequences are introduced into cells in culture using techniques known in the art that include but are not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake. Treated cells are monitored for altered cell function. These cell functions are predicted based upon the homologies of the gene, corresponding to the EST from which the oligonucleotide was derived, with

known genes sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the EST is associated with the disease using techniques described in Example 20.

EXAMPLE 23

Gene expression from DNA Sequences Corresponding to ESTs

A gene sequence of the present invention coding for all or part of a human gene product is introduced into an expression vector using conventional technology. (Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art.) Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield, et al., U.S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide from cloned cDNA sequences. The cDNA from the EST of interest is sequenced to identify the methionine initiation codon for the gene and the poly A sequence. If the cDNA lacks a poly A sequence, this sequence can be added to the construct by, for example, splicing out the Poly A sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct

allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The cDNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the cDNA and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the cDNA is positioned inframe with the poly A sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferrably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface.

Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted cDNA sequence are injected into mice to generate antibody to the polypeptide encoded by the cDNA.

If antibody production is not possible, the cDNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as a chimeric with, for example, β-globin. Antibody to β-globin is used to purify the chimeric. Corresponding protease cleavage sites engineered between the β-globin gene and the cDNA are then used to separate the two polypeptide fragments from one another after translation. One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene). This vector encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal

incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al. and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from either construct using in vitro translation systems such as In vitro Express™ Translation Kit (Stratagene).

EXAMPLE 24

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 23. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., **Nature 256**:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid

of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, E., **Meth. Enzymol. 70**:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. **Basic Methods in Molecular Biology** Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. **J. Clin. Endocrinol. Metab.** 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: **Handbook of Experimental Immunology** D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: **Manual of Clinical Immunology**, 2d Ed. (Rose and Friedman, eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample.

EXAMPLE 25

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Example 24 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunchistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: **Basic & Clinical Immunology**, 3rd Ed. Lange, Los Altos, California (1980) or Rose, N. et al., Chap. 11 in: **Methods in Immunodiagnosis**, 2d Ed. John Wiley & Sons, New York (1980).

A fluorescent marker, either fluorescein or rhodamine,

is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single gene copy or protein, identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, L. et al., Section 19-2 in: **Basic Methods in Molecular Biology** (P. Leder, ed), Elsevier, New York (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5-50 µl, and containing from about 1 to 100 µg protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. et al., Basic Section 19-3. One set of

nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Example 24. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from EST sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

The entire contents of all references cited above are hereby incorporated by reference.

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

VII. Correlation of EST and Clone Identifiers

The EST sequences of the present invention are identified herein by SEQ ID NO, and are identified in the GenBank

database by a different number, are identified in the inventors' lab (and upcoming publications) by EST number, and clones have been submitted to the American Type Culture Collection (Rockville, Maryland USA) under clone names. Table 12 cross references those different numbers for the ESTs from cDNA, SEQ ID NOS 1-315.

Table 12. SEQ ID NO Cross References

SEQ ID	EST#	GB#	Clone
1	EST00007	M61959	HFBA01
2	EST00009	M61953	HFBA05
3	EST00010	M61961	HFBA07
4	EST00011	M61962	HFBA08
5	EST00012	M61963	HFBA10
6	EST00013	M61964	HFBA11
7	EST00014	M61965	HFBA11
8	EST00234	M62172	HFBA26
9	EST00376	M61966	HFBA20
10	EST00016	M61967	HFBA23
11	EST00018	M61968	HFBA36
12	EST00274	M62212	HFBA51
13	EST00055	M62194	HFBA65
14	EST00019	M61969	HFBA66
15	EST00020	M61970	HFBA69
16	EST00021	M61971	HFBA71
17	EST00022	M61972	HFBA77
18	EST00373	M62299	HFBA84
19	EST00023	M61973	HFBA86
20	EST00024	M61974	HFBA87
21	EST00025	M61975	HTBA89
22	EST00301	M62239	HFBA90
23	EST00026	M61976	HHCA08
24	EST00027	M61977	HHCA05
25	EST00028	M61978	HHCA08
26	EST00310	M62245	HHCA02
27	EST00029	M61979	HHCA08
28	EST00030	M61980	HHCA09
29	EST00031	M61981	HHCA104
30	EST00032	M61982	HHCA10
31	EST00033	M61983	HHCA11
32	EST00233	M62171	HHCA11
33	EST00034	M61984	HHCA13
34	EST00035	M61985	HHCA14
35	EST00036	M61986	HHCA18
36	EST00037	M61987	HHCA21
37	EST00038	M61988	HHCA23
38	EST00374	M62300	HHCA53
39	EST00039	M61989	HHCA53
40	EST00040	M61990	HHCA54
41	EST00041	M61991	HHCA54
42	EST00042	M61992	HHCA55
43	EST00371	M62297	HHCA57
44	EST00364	M62290	HHCA68
45	EST00044	M61994	HHCA69
46	EST00046	M61995	HHCA71
47	EST00291	M62229	HHCA73
48	EST00047	M61996	HHCA79
49	EST00048	M61997	HHCB04
50	EST00049	M61998	HHCB08
51	EST00052	M61999	HHCB16
52	EST00054	M62000	HHCB35
53	EST00055	M62001	HHCB51
54	EST00056	M62002	HHCB52
55	EST00057	M62003	HHCB53
56	EST00058	M62004	HHCB55
57	EST00059	M62005	HHCB96
58	EST00061	M62006	HHCC04
59	EST00062	M62007	HHCC05
60	EST00290	M62228	HHCC07
61	EST00064	M62008	HHCC09

SEQ ID	EST#	GB#	Clone
63	EST00065	M62009	HHCC10
64	EST00066	M62010	HHCC13
66	EST00067	M62011	HHCC18
67	EST00351	M62200	HHCC21
68	EST00068	M62012	HHCC22
69	EST00360	M62207	HHCC23
71	EST00070	M62014	HHCC27
72	EST00071	M62015	HHCC29
73	EST00072	M62016	HHCC31
74	EST00073	M62017	HHCC37
75	EST00074	M62018	HHCC40
76	EST00075	M62019	HHCC42
77	EST00257	M62196	HHCC53
78	EST00258	M62197	HHCC53
79	EST00076	M62020	HHCC64
80	EST00077	M62021	HHCC67
81	EST00315	M62248	HHCC70
82	EST00078	M62022	HHCC72
83	EST00079	M62023	HHCC74
84	EST00080	M62024	HHCC76
85	EST00081	M62025	HHCC77
86	EST00082	M62026	HHCC78
87	EST00083	M62027	HHCC80
88	EST00084	M62028	HHCD01
89	EST00085	M62029	HHCD02
90	EST00302	M62240	HHCD03
91	EST00086	M62030	HHCD05
92	EST00087	M62031	HHCD06
93	EST00287	M62225	HHCD08
94	EST00353	M62281	HHCD10
95	EST00088	M62032	HHCD12
96	EST00089	M62033	HHCD14
97	EST00289	M62227	HHCD17
98	EST00260	M62199	HHCD21
99	EST00316	M62249	HHCD30
100	EST00090	M62034	HHCE04
101	EST00091	M62035	HHCE05
102	EST00248	M62187	HHCE06
103	EST00317	M62250	HHCE07
104	EST00354	M62282	HHCE08
105	EST00365	M62291	HHCE10
106	EST00092	M62036	HHCE12
107	EST00093	M62037	HHCE13
108	EST00094	M62038	HHCE15
109	EST00095	M62039	HHCE17
110	EST00096	M62040	HHCE20
111	EST00281	M62218	HHCG01
112	EST00318	M62251	HHCG03
113	EST00097	M62041	HHCG07
114	EST00098	M62042	HHCG10
115	EST00099	M62043	HHCG14
116	EST00100	M62044	HHCG18
117	EST00319	M62252	HHCG25
118	EST00101	M62045	HHCG29
119	EST00102	M62046	HHCG31
120	EST00103	M62047	HHCG33
121	EST00104	M62048	HHCG36
122	EST00105	M62049	HHCG37
123	EST00106	M62050	HHCG38
124	EST00107	M62051	HHCG40
125	EST00108	M62052	HHCG41

SEQ ID	EST#	GB#	Clone
126	EST00109	M62053	HHCG44
127	EST00320	M62253	HHCG51
128	EST00252	M62191	HHCG57
129	EST00321	M62254	HHCG60
130	EST00355	M62283	HHCG61
131	EST00322	M62255	HHCG63
132	EST00110	M62054	HHCG65
133	EST00111	M62055	HHCG66
134	EST00375	M62186	HHCG67
135	EST00112	M62056	HHCG68
136	EST00113	M62057	HHCG70
137	EST00272	M62210	HHCG72
138	EST00114	M62058	HHCG85
139	EST00116	M62059	HHCH01
140	EST00117	M62060	HHCH09
141	EST00118	M62061	HHCH11
142	EST00323	M62256	HHCH15
143	EST00119	M62062	HHCH36
144	EST00120	M62063	HHCH41
145	EST00121	M62064	HHCH49
146	EST00122	M62065	HHCH50
147	EST00292	M62230	HHCH51
148	EST00236	M62174	HHCH52
149	EST00123	M62066	HHCH53
150	EST00124	M62067	HHCH57
151	EST00125	M62068	HHCH59
152	EST00126	M62069	HHCH60
153	EST00127	M62070	HHCH62
154	EST00128	M62071	HHCH63
155	EST00129	M62072	HHCH64
156	EST00130	M62073	HHCH66
157	EST00131	M62074	HHCH68
158	EST00132	M62075	HHCH71
159	EST00325	M62257	HHCH72
160	EST00326	M62258	HHCH73
161	EST00247	M62185	HHCH76
162	EST00133	M62076	HHCH77
163	EST00134	M62077	HHCH79
164	EST00135	M62078	HHCH83
165	EST00136	M62079	HHCH84
166	EST00137	M62080	HHCH87
167	EST00138	M62081	HHCH91
168	EST00140	M62082	HHCH96
169	EST00141	M62083	HHC105
170	EST00295	M62233	HHC106
171	EST00327	M62259	HHC110
172	EST00142	M62084	HHC117
173	EST00143	M62085	HHC126
174	EST00296	M62234	HHC127
175	EST00144	M62086	HHC133
176	EST00328	M62260	HHC136
177	EST00294	M62232	HHC142
178	EST00145	M62087	HHC147
179	EST00299	M62237	HHC154
180	EST00147	M62088	HHC155
181			

SEQ ID	EST#	GB#	Clone
182	EST00329	M62261	HHCI59
183	EST00148	M62089	HHCI61
185	EST00149	M62090	HHCI62
186	EST00150	M62091	HHCI73
187	EST00151	M62092	HHCI75
188	EST00152	M62093	HHCI79
189	EST00256	M62195	HHCI84
189	EST00282	M62219	HHCI85
190	EST00153	M62094	HHCI86
191	EST00154	M62095	HHCI88
192	EST00155	M62096	HHCI90
193	EST00156	M62097	HHCI92
194	EST00157	M62098	HHCI93
195	EST00158	M62099	HHCI94
196	EST00159	M62100	HHCJ05
197	EST00160	M62101	HHCJ07
198	EST00161	M62102	HHCJ09
199	EST00277	M62214	HHCJ13
200	EST00162	M62103	HHCJ15
201	EST00163	M62104	HHCJ17
202	EST00298	M62236	HHCJ29
203	EST00164	M62105	HHCJ30
204	EST00235	M62173	HHCJ34
205	EST00165	M62106	HHCJ35
206	EST00166	M62107	HHCJ36
207	EST00167	M62108	HHCJ37
208	EST00250	M62189	HHCJ42
209	EST00331	M62262	HHCJ43
210	EST00168	M62109	HHCJ47
211	EST00332	M62263	HHCJ48
212	EST00169	M62110	HHCJ50
213	EST00170	M62111	HHCJ51
214	EST00171	M62112	HHCJ59
215	EST00172	M62113	HHCJ60
216	EST00173	M62114	HHCJ61
218	EST00175	M62116	HHCJ67
219	EST00176	M62117	HHCJ73
220	EST00372	M62228	HHCJ74
221	EST00359	M62286	HHCJ78
222	EST00177	M62118	HHCJ79
223	EST00369	M62294	HHCJ80
224	EST00356	M62284	HHCJ85
225	EST00178	M62119	HHCJ84
226	EST00333	M62264	HHCJ85
227	EST00259	M62198	HHCJ86
228	EST00179	M62120	HHCJ91
229	EST00180	M62121	HHCJ89
230	EST00181	M62122	HHCJ92
231	EST00334	M62265	HHCJ93
232	EST00182	M62123	HHCJ96
233	EST00183	M62124	HHCSA26
234	EST00184	M62125	HHCSA28
235	EST00185	M62126	HHCSA29
236	EST00186	M62127	HHCSA30
237	EST00187	M62128	HHCSA31
238	EST00188	M62129	HHCSA49
239	EST00189	M62130	HHCSA53
240	EST00335	M62266	HHCSA59
241	EST00191	M62131	HHCSA64
242	EST00192	M62132	HHCSA70
243	EST00193	M62133	HHCSA75
244	EST00194	M62134	HHCSA77
245	EST00347	M62275	HHCSA79
246	EST00196	M62135	HHCSA81

SEQ ID	EST#	GB#	Clone
247	EST00279	M62217	HHCSA82
248	EST00271	M62209	HHCSA85
249	EST00275	M62213	HHCSA86
250	EST00197	M62136	HHCSA87
251	EST00370	M62296	HHCSA92
252	EST00198	M62137	HHCSA95
253	EST00199	M62138	HHCSB05
254	EST00200	M62139	HHCSB07
255	EST00201	M62140	HHCSB08
256	EST00345	M62267	HHCSB09
257	EST00337	M62268	HHCSB12
258	EST00346	M62276	HHCSB13
259	EST00202	M62141	HHCSB19
260	EST00357	M62285	HHCSB20
261	EST00338	M62269	HHCSB21
262	EST00339	M62270	HHCSB22
263	EST00203	M62142	HHCSB23
264	EST00204	M62143	HHCSB25
265	EST00205	M62144	HHCSB27
266	EST00206	M62145	HHCSB29
267	EST00297	M62235	HHCSB30
268	EST00369	M62295	HHCSB31
269	EST00293	M62231	HHCSB32
270	EST00207	M62146	HHCSB34
271	EST00283	M62220	HHCSB37
272	EST00340	M62271	HHCSB43
273	EST00208	M62147	HHCSB44
274	EST00268	M62206	HHCSB45
275	EST00209	M62148	HHCSB46
276	EST00211	M62149	HHCSB48
277	EST00212	M62150	HHCSB50
278	EST00342	M62272	HHCSB53
279	EST00213	M62151	HHCSB55
280	EST00343	M62273	HHCSB56
281	EST00214	M62152	HHCSB57
282	EST00344	M62274	HHCSB58
283	EST00215	M62153	HHCSB62
284	EST00216	M62154	HHCSB68
285	EST00286	M62224	HHCSB69
286	EST00217	M62155	HHCSB70
287	EST00218	M62156	HHCSB73
288	EST00219	M62157	HHCSB77
289	EST00220	M62158	HHCSB78
290	EST00221	M62159	HHCSB79
291	EST00222	M62160	HHCSB81
292	EST00223	M62161	HHCSB83
293	EST00224	M62162	HHCSB84
294	EST00225	M62163	HHCSB91
295	EST00226	M62164	HHCSB92
296	EST00228	M62166	HTCA08
297	EST00230	M62168	HTCA27
298	EST00231	M62169	HTCA27
299	EST00249	M62188	HTCA36
300	EST00232	M62170	HTCB03
301	EST00300	M62238	HTCB05
302	EST00303	M62241	HFBA22
303	EST00348	M62277	HFBA45
304	EST00307	M62242	HFBA78
305	EST00308	M62243	HHCA05
306	EST00309	M62244	HHCA06
307	EST00312	M62246	HHCA57
308	EST00314	M62247	HHCA63
309	EST00174	M62115	HHCJ67
310	EST00377		HHCG05

SEQ ID	EST#	GB#	Clone
311	EST00270	M62208	HFBA18
313	EST00276	M62215	HHCJ13
315	EST00008	M61960	HFBA04

NOTE REGARDING SEQUENCE LISTINGS: The listings of SEQ ID NOS: 1-315 are in numerical order. However, an occasional number (for example, SEQ ID NO: 44) is not found in this list. In all, 7 SEQ ID NOS are not used. Nevertheless, the convention "1-315" is used, for example, to refer to all the SEQ ID NOS in the following list.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Venter, J. Craig
Adams, Mark D.
Moreno, Ruben F.

(ii) TITLE OF INVENTION: Sequences Characteristic of Human Gene Transcription Product

(iii) NUMBER OF SEQUENCES: 308 (1-315, with 7 SEQ ID NOS unused.)

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Knobbe, Martens, Olson, and Bear
(B) STREET: 620 Newport Center Dr. Sixteenth Floor
(C) CITY: Newport Beach
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92660

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/716,831
(B) FILING DATE: 20-JUN-1991

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Israelsen, Ned A.
(B) REGISTRATION NUMBER: 29,655
(C) REFERENCE/DOCKET NUMBER: NIH004.004CP1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 619-235-8550
(B) TELEFAX: 619-235-0176

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 362 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATCCTAGTTG ATGGTCTGGG TTATCAGAGG AGCAAAAACA TTTAAGTGTC AAATAATGCT 120

CATTGTCTCC CTGGGATTTC TAAACAGAAA AAATGAAGAA AGAGGCAGAG AAGAGCTTCA 180

CAAGGTGTGT GCCAGCTCTG CATCATTTCC AGCTGCTCAA CCACCATTTC TCCCATTTTA 240

GGTCCCCAAA AGTAGGAGGT GGGGCCTCAC AGAGCTGCTG TGGGCTTTGG GTATCAAAAG 300

CTGCAGCCAC CATATGGGGC ACTCCTGGCT GGTGTACAGG GTGGGCATTG CCCAGGTCTT 360

TT 362

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTTTTNCTTT TTTCTTAGCT TCATTTCTCT TAAAAAACAA GGAACAAGAA AACATTGCAC 60

CAGCGTTCTA AGCCTCAAAC AAAANACAAA ACAAATCCCC CTGCGAAGAA CAATAAACTT 120

TACATCTCTT TGGCAACAAT AACTTAAAAT CACCCAACTT CCATTCGCTC CAACCACAGC 180

AGTTAGTTAG TTACAAAAAT ATTCCNTGTG CTGC 214

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATTAATAGGA AAGATGATTG TATAGATGGT GGGCTATTAA CTCAGATCAG GATGAGAATC 60

GGGAGTGCCT TTACATGTGT GGTACCCAAA TGGGTGGTTG GATATAAGAG TAACAAAAGG 120

ACTGAAAGGG TTAAAAAAGA AAGAAAAAAA AAAAACTCCC TGGTTGGGAG GGTGTTAAGT 180

ATCGAGTGTT TTTCCAAACC ATTCCTCCTC TGCTCACCTA CCCCTAGGTG ATTAAAGGAG 240

ATAACTTTTA AAAAAGAAAG AATTGGCTCA AAGGTACTGT AAATTCTAGG ATTATATACC 300

TTTATATAGG TTCATTCCCT GATCCCTGTA TTATCAAGGC ACAG 344

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGACACTCAT CTGTGCCCCA CGTCCAGATC CCTGGAGGCA GCTGACCAAT GATGGGCGGT 60

GACCCGGTAA CCGAGGCGGC AAGGAGGCCA GGTAGTCCCG GCACCTCTCA CTCTGCAGAG 120

ACCAGCGGCT TCGTGGGAGG CCTGTGGGTC ACACGTAGGG GCTAGAGCCA GCCTGCATCC 180

TGCCCACCGG GCTCCACTTG GAGATCAGCA GGAGGGCCAG TGTGGGACCC CTGCTGCCAC 240

CTCTCCTGGG CCTGTKTCCT TTCTGGAAAT TAAGAAGGTG TGCTCCAGAG CCAAGAGGAG 300

CAATAAGAAA CCTCGTGTGC CAGCTTCTTA AGGGTKGCAG TGCAAGACCC CA 352

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 562 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATAGCCTTAC ATATATATTC ACAGAAAATC ATATTGCATA TACTCTTTCT CCACATCATA 60

AAAATGGGTG TTGGGCTCTC TAGGACACAA GGGAAGCAGG CCAAATTTCT CATATTTTCA 120

GGAATAAACT GAGTGCCCCG AAGGTGTAAT AGGAACCTTT TACTAACCTC ATCTGACTTC 180

ATCCTCACAC CAGCATTTTG TGTGTAAGGA AACTGGCCGA GACTGGTTAA GAAATATATC 240

CAAAGACGTA TACTTCCAAA TGGAACACGG ATCTTTTTAT TTAAATTCCA ATCATCTTTC 300

CATTATATCA GCCAATGATG GAGCAGAAAG CTGGTCCAGG CAATCCCAGA ATAGATCTTT 360

CTAGCCACCC GTTCAGTGTG AGGAGGGGGA AGTGGCCTTG CCAAGGGGCC AGTGAGCTCA 420

ATTAGGGTTA ACGCTGCTTC TTAGCCTACC CCAGGGGNCA CCGCACTTAG GTTGTTTTGT 480

GCCCAGCTTT GGCAGGAAGC ATTCCTCCTT TCAAAGATTN NAGCCTTGCC GTCATATATC 540

GGGTGTAATA GGGTTCTTTT TT 562

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGATCTTCTC CCTCTTTCAA TTTTAGCATT AATGTGATCC TCAAAAAATCC ATTAATACTA 60

GTTAAACTAA ATAAAGCCAA GAGCTCCAAA ATCCCTCCAT TAAATCCATT TTTCCACACT 120

AATCTCAATC ATCCAAAGCT ATTTTCAATA AGTCAGCTAT TCAAAGCTAT TCACACTACT 180

TGAAAGAGTA ATTACCATTT ACTGAAGCAC TTATCTGTCC TACACTGATG GGAGTAAATG 240

CTTCTCATAG GTTATCTCAT GTACATTATG CCACTTTNAC TTAAAATGAT CACAATTNAG 300

TGCTATAGGT TTTTGGGTTA ATGTTTTCCC NGGGGGAGTT GTTAAAAACA TGGCATTTC 359

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AACTTGCAAC ATAAATACTA GAAAAAGAGA AAATATCATC AAAATACAAA TAACTGTTAG 60

AAATCATTGC TCAAAAGAAR AACCTGGCAA TGCATGATTA CGAAATGCAA AAGAMGATAC 120

AGTTGCTCTC TGTATATGCG CTTTCCACAT CCACAGATTC AAACAACTGT GGATAAAAAA 180

GGATTTTTCA ATGCCATTAA ACAVCAATGC AACAGTAA 218

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTACAATAGA AGGCAAACTA TGTCCCTCCT TTGCTCAGAA ACTTTTAATA TCTKCCTATT 60

TCCCCATGTA AAAGCCAATC CTCAACCACA GTGTAGAAGG GCTATCCATT TCTAGCTACA 120

CATCTCCTCA GTCACTGCCC CCAGCCCCAG TACTTGGGGA CTTTGCCCTT GCAGTTCCCT 180

GTGCCAGCAA ACTCTTCCTC CAGATGTCCA CATGACTCAC CCNNCTCCTT CAGGGGTCTT 240

CTCAAATGTC ACTTTACCAG AGGTGGCTTC CCTGACCATC CTGTATAAAT AGCATCACCC 300

TACCTCCTAT CTCTCTCTCT AATGTCTCAG GAATTCGATA TCAAG 345

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 189 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTGAACAGAC TAAGGCCTTT NTGGAGGCCC AGAATAAGAT TACTGTGCCA TTTCTTGAGC 60

AGTGTCCCAT CAGAGGTTTA TACAAAGAGA GAATGACTGA ACTATATGAT TATCCCANGT 120

ATAGTTGCCA CTTCAAGAAA GGAGAACGGT GTTTTTATTT TTACAATACA GGNTTTNAGA 180

ACCACCGGG 189

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTGCCTTCGC CACCTGCTGG ACGCGAGGGG CTACTACGAT GCCATGGGTG TCCTGRTTTT 60

TTATTTCTCA GACAGGACTG CTCTGTATNT GTCTTTGGAT TCTACGTAGA TTTATATTTG 120

TAAAATATTA CATTTGTCAT GACCAGAAGA AATGTCATTA TCGTAAAATT TAGATTCTGG 180

NGTCTATATA TGNAAGNAAT ACTAACTACT AACTGTTATA ACAWCAAAAT GTGGGNTGTA 240

TATCTACARG CCNGAGCCGA CTTGTCA 267

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 247 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTCATAAAGC CAGGGTGATA AAAWTGGTAG TTTCATGTTA TCTACAAGRC TAAGKTCAAA 60

ATTCCATGCA TGTGCTGRTA AAAGACCCAT NATGGKCCTM ACTGTACTTA CTCCCCATTT 120

ATTAGCATTC ATTCTGGTCA CCAGCTCTAG TTCCTCTGCT TAGCGAATCT CGCTTGTCTT 180

CAAGATGTCA TTCAAAATGTC ACATTTTGTG GGAAGCCTTG CCTTTTTTGA CACGGTCTCC 240

CTGCCAC 247

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 280 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AAGCCCACAG GCTTCTGGAC AAACCCACCC CATCAACCTC TTCATCTTGG ACTTTTANCC 60

TCCAGAGCTA TGAGAAAACA ANTTTCTGTN NATTGNGGGC ACTCAGCCTC TGGATACTGG 120

CAGCCCTACC AAACTCATAC ACACATACAT TTTAAATTC TTTAATTCCT GTGACCATTC 180

ACTTATGGTT CACTTTTTAA ATACTCCTAC TCTTATTTTT CTTGTTAAAG TTCACCACGA 240

CATAGCTCAT TGCCCAAAGC CCCTTCTAAC TTTT 274

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

VCTVTCTVCC AACTTCATTC AGATATTGAC TCTGGTGATG GGAACATTAA ATACATTCTC 60

TCAGGGGAAG GAGCTGGAAC CATTTTTVTR ATTGATGACA AATCAGGGAA CATTCATGCC 120

ACCAAGACGT TGGATCGAGA AGAGAGAGCC CAGTACACGT TGATGGCTCA GGCGGTGGAC 180

AGGGACACCA ATCGGCCACT GGAGCCACCG TCGGAATTCA TTKTCAAGGK CCAGGACATT 240

AATGACAGTC CTCCGGAGGT TTCCTGCACG AGACCTATCA TGCCAACTGT GCCSTGTARA 300

GGTCCAATKT TGGGTGSTGT ACGGTAGTGG GGAGGCCTG 339

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGGVGCAAAG TAGCAGATTC TAGTAAAGGA CCAGATGAGG CAAAAATTAA GGCACTCTTG 60

GAAAGAACAG GCTACACACT TGATGTGACC ACTGGACAGA GGAAGTATGG AGGACCACCT 120

CCAGATTCCG TTTATYCAGG TCAGCAGCCT TCTGTTGGCA CTGAGATATT TGTGGGAAAG 180

ATCCCAAGAG ATCTATTTTG AGGATGAACT TGTTCCATTA TTTGAGAAAG CTTGGACCTA 240

TATGGGATCC TTCGTCTAAT GATGGATCCA CTCACTGGTC TCAATAGAGG TTAATGCGTT 300

TGTCACTTTT TTGTACAAAA GGAGCARGCT CAAGGAGGGC TG 342

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATGTTGATGC TGAAATTVAA GATCCACCAA TTCCAGAAAA ACCATGGAAG GTTCATGTGA 60

AATGGATTTT GGACACTGAT ATTTTCAATG AATGGATGAA TGAGGAGGAT TATRAGGTGG 120

ATGAAAATAG GAAGCCTGTR AGTTTYCGTC AGCCGATTTC AACCAAGAAT GAAGAGCCAG 180

TCAGAAGTCC AGAAAGAAGA GATAGAAAAG CATCASCTAA TGCTCGAAAG AGGAAACATT 240

GGCGTTCGCC TCCCCCTCGG ACACCAACAG AWTCACGGGA AGAAGAGTGG GAAGAAAGGC 300

CAAGCTAGCC TTTTATGGGG AAGCCGCAAG AAGTCCAGAA AGAGGGWWGG TTGA 354

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGGCAAGTT TCTTCCAGGA TGAGAAATCA GTGGAAAGTG AGGGCCAGCC AACAGCCACC 60

ACCAACCACC CAACACGCGA GCGAGACCAT CTTAAAAGAG CCCCAGCCAA GCTGACCATG 120

GGTCTGACCC CAAACTGAAG AAATGCCCAG CCCAGCCAAA CCCAAATTGC TAACTTGTAT 180

TATAAGCAAG TACAATGGTC CTTACCTTAA GCCACTAAGT TTTGGGATGC TTTGTTACAC 240

AGCTATAGAT AAGCTGATAC AGGGAATGTC AGAWTCCATG ATGAGAGACC GAGCCTTTCA 300

NTCTGTCAGA GGYACCTTVG GTTGGCAAAA CTTCAAAAAG AGGGACCT 348

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGCATGGGCT GGGGGGCCGG GACTTAGGGC TGGGGCTTGT TTTAGGCTCT GCCCCCCACA 60

CCCCCTCCTC TTCCGTCCTG ATTAAGCCCA AGGGTTGGTG GACTTAACTT TCAGCCCATC 120

TCTAAGGGTT TCACAGACTG GATCTTTCTA AACTTTATTG GGTACCTGCT TCCCCTTTTC 180

CCTGGTAGTT TTGATCTACA AAAAGTCAAA ACCTGATCGA AATAGAAATA AGATCATCAA 240

ATTGGACCAT TCTCTTAGCG TTCGAGTGTG CCGGCCAGAC TGGCATTCAC TACACGCTGA 300

GATCCAACCA CATCACACTG GCCTCAGGTC ACCAACTCGC CACTCAGGGC ACAAGCCCTG 360

CCCTTGTGGT CACAAGGCTT TCCTTAATGT CGTCGGTGCC CAGGTGAACC ACAAC 415

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATATCTATCT CTGTAGCTAT TTCTACACTT AACGATCTGT GTCCTAATTA AGCTAAACAT 60

GATTCATTCT GATGCCAACC CCCATCCATC ATGGCATGGA TCGCTCTAGA CTTCTTCCCT 120

TGTAACCTCC CACTCAAACA GTGAGAAACC TTTGCCCAGT ATGTTTTGGA GTAACCTCAC 180

TGGGAGTTTG CAGTCCCACT AGATGAATGC CAACCCATTT GTTCATTTAA AAGGACTTTT 240

GGAACCATAG AGCAATGGCT GGGCTGGGTC TVGCACGTTC ATCTTGACTG AAACAATTGG 300

CCATGAAGGC ACTTGCCAAG GAAACTCTAG GGGCCACAAG GGTCCTGGGT GCTTGC 356

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CATGCTTCCA TTTTTTTTAG TTTTAAACCA CCAAACCAAT ATTTTYCCTT TAAATTTTAA 60

TCTTATAATA TAGAAATCTT ATGTAAATGA AATTTTGTCA TGTTTCAAAT AAAGAGAACT 120

GAAGTAGAAA ATAGAAATGC CAGTAAACAA CATAATGTTT AATTTACAAC TTACATTAGG 180

GGTTTGGGGG VATGCTAATT ATATATTGAG AATATACATT AGAACTCTTC AAAATGGGCT 240

CTTCTAATGA GGTCACTACT GAACATAATT GTTCCCTCTT CTGTTAAATA GAATAGGTTT 300

AAATGACTAG TCCAAATGGA ATTATTGCCT TCTKGTTAA 339

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGAACAAGGG AACTCAGCAG CCCCTCCCTT CCCATCAGCT GTTCCTGAGA GATGCAATAT 60

AGTAGTCATC GACATCATCC TTATCAACAG CATCATCACT CAGACAGTGG TGAAAGTCTT 120

TCTTCACAAG GAAAAACAAA GATAAAGAAA TACATGAGCA TTAATCAGAA ATTTTCAAAG 180

CTTGGATTCT AATGATATGC ATTATCATTA GACATTCAAA TGCTATACAT CTTCTGATGA 240

AGCCTCCTTG ACAGCAGCTA CACTTATTTC ACATTAGAAT GCCTAGAGAA ATCCTGACTG 300

CCCAGCTTGG TCATGGGACC TTCCCCACTC TCCTCTTGGA GGAATGAAAA GATGTGGCGG 360

CTTTCTACTT TTGCTACTGA GCTGGGGTAT ATGGCTAGGT CCACTTTCTA AGGGGCTTGG 420

AAGGGTTATT CCATCTG 437

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 385 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTTTGATTTG CTTTTTTTTT AGAGTTTTAC ATCAGTGTTT TTCAGGAATA TTGGTCTTTC 60

ATTTTCTTTT CTTGGAATAT TTTCTAGTTT TACTTTGTCA GAGTAAATTC TGGCTTCACA 120

GAATTATTTG TAGTCTCTCC TGTCTTGGTT TATTCATGCT GCTATAACAA AATACCACAG 180

ACAAGGTGGT AATAAATAAC ACAAATTTAT TTTTCCCAGT TCTGGAGGCT AGGAGTTCAA 240

GAAGCTGGCA AGTTCAATGT CTGGTGAGAC CCATTCCTTC ATAGGTGGCA CCATCTAGGG 300

GTCCTTACAT GRCAAAGAGA TGGAAGGCCC AAAAAGATGG TGACCTATTG TGAGGCCTTT 360

TTTAAAGGGC CTTVAAATCC CAGTC 385

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ACCTTCATGG TCATGAAGGC CATGCAGTCT CTCAAGTCCC GAGGCTACGT GAAGGAACAG 60

TTTGCCTGGA GACATTTCTA CTGGTACCTT ACCAATGAGG GTATCCAGTA TCTCCGTGAT 120

TACCTTCATC TGCCCCGGGA GATTGTGCCT GCCACCCTAC GCCGTAGCCG TCCAGAGACT 180

GGCAGGCCTC GGCCTAAAGG TCTGGGAGGG TGAGCGACCT GCGAGACTCA CAAGACGGGA 240

AGCTGACAAG AGATACCTAC AAGACGGGAG TRCCTGTGCC ACCTGGTGCC GACAAGAAAG 300

CGGAGGCTTG GGTCTGGGTC AGCAACCGAA TTCCAGTTTA GAGGCGGATT TVGGTCGTKG 360

ACGGTGTCAG CCAC 374

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 322 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GAAAACGTGA TGACGACAGC TCCGTTCGTG TAGTGACACT TAACATACTC ACCATCTTCA 60

TGAATTCTGA ATAATTTACT GATCGTAAAG TCTAAAACGA TGAATTTGAG GTGACCAGTT 120

TTAAATCAGA AAATAGTCAA TAGTTAATCA TGACTCTTCA GCGTATTTTC TTTACCTCCT 180

CTGAAGAGTT TCCCAGAACA TTCTTGTGAA AAGGAATGCC TCCCAACAAT GGAGAGCAAC 240

AATAGCAACA GGCATCTGAA TCAGCCTGGC CTCTGAAAAC AGACCANAGA GGAGTTTATC 300

TGTTTCTTCC AGTGGAGGAA GG 322

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCTGAAATCG GAGTCTTTTG GACTGACTCC AAATTCAATG GGTGGCACAG GCAGCACGGA 60

GTCCAGGTGA ATCTCCACCC CGTTAACAGG CGGGACGACA GCCCCTTGCA GCC 113

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGAAAGAATG AAGGAAAAAC AAGACAAAAT CTACTTCATG GCTGGGTCCA GCAGAAAAGA 60

GCAGACGCTG GCCTCAGACA CAGACAGCAG TCTTGATGCC TCGACGGGAC CCCTTGAAGG 120

CTGTCGATGA TAGGTTAGAA ATAGCAAACC TGTCAGCATT GAAGGAACTC TCACCTCCGT 180

GGGCCTGAAA TGCTTGGGAG TTGATGGAAC CAAATAGAAA AACTCCATGT TCTGCATGTA 240

AGAAACACAA TGCCTTGCCC TACTCAGACC TGATAGGATT GCCTGCTTAG ATGATAAAAT 300

GAGGCAGAAT ATGTCTTGAA GAAAAAANTT GCAAGCCACA CTTCTNGAGA TTTTGTTCAA 360

GATCCATTTC AGGGTGAGCA GTTAGAGTAG GTTGAATTT 399

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GATTGGTATA CGGGCAACAA TGGATTGATA GCCTTAATAT AGAAATAGTT CCAGCAGGCC 60

AGATGCAGTG GCTCAATTCT GTAAACCCAG TGCTCTGCAC AGCTAGGAAG GAAGATCACT 120

TGGGCCCAGG AGTTCAAGGC TCCAGTGAGC CATGATCACG CCACTKCCTC CAGCCTGGGT 180

GACAGAGTNA GGCCCTGTCT CTAAAAAATG AAATAGCTCC ATCAAGTCAA TAATTAAAAG 240

TTCAACAGCC CAACAGANCA AAAATTGTAA ATGANCACAA ATTAGAAAAT GTACAAATTA 300

AATATTAATG ACCCATAACC CTATAAGGGA AAGTTTAACC TCTCTAGTAT TTTTT 355

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 322 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AAAACGTGAT CACCACAGCT CCGTTCCTGC AGTGACACTT AACATACTCA GCATCTTCAT 60

GAATTCTGAA TAATTTACTG ATCGTAAAGT CTAAAAGTAT CAATTTCAGG TGAGCAGTTT 120

TAAATCAGAA AATAGTCAAT AGTTAATCAT GACTCTTCAG GGTATTTCCT TCACGTCCTC 180

TGAAGAGTTT CCCAGAACAT TCTTGTGAAA AGGAATGCCT CCCAACAATG GAGGAGCAAC 240

AATAGCAACA GGCATCTGAA TCAGCCTGGG CTCTGAAAAC AGACCAAAGA GGNGTTTTTC 300

TGCTTTCTTC CAGTGAGGAA GG 322

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 287 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TATTTTTATT AAAGGACCAC CCTGGCTGTM GTGAGATGAA TGGATTCAAA CAGGGCAAGA 60

GTGGATACAC MGAGATAAGT TAGGAAGCTG GTATAGAAAT CTGGATGAGA TATGGTGGCT 120

TGGATGATAC TAGCAGTGAG TATGGGAAGT AGGTGGATTA CTTTACACTT TTTTAGATCA 180

GTCNATTCTT GATGTCTTGA AGACAAATTA ATCTCATATA TAACTCTAAA CAACATATTT 240

ATATTTCATG TAAATAAGGA TAATGCTGAC CAAATATTAG CACCTTT 287

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAGGGCAGGG AAGCCTGGAA GCAAAGGAGG AGGTGGGTGC TGACTCTGAG AGAGGATAGG 60

CTGGGATCCC TAGGGCAGGC TTGTTCCTCG GGTGGCCAAT TTAGTCTTTT AATTCTCTAA 120

GGGCTCTCCA TTGCCTGCCC TTGCCTCTTT CTAGCCTGTT ATTTCTAGGC TCCTCTGAAT 180

AAATCTCAGG TTTCCTACTG TCATGCCTTT AGTTCAAAAA TGAGAATCTG CCCTACAGTG 240

CTGGCCTCCT TCCGGCCTGA AAGCCAGCAC CTTKCGACCC GG 282

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAAGCTGGTG AATACATTTC AAGACACAAC ATGGCACCTG TGTCTAGCTC TATGGTACAA 60

CATGGTACTA TGACACATAT AATGGGTTGC CAGATGGGGA AGGCAGCTTC TCTGCAACTG 120

AGCTGAGATC TCAAAATAGA CAATGTCAAG ATGGAATGAG AAGGGAAAAA CAGCATGTGT 180

AGACAGGTAG TGACAAAAGG CTAATTAAGG ACTGAAAGAA ACCAGTGGCC AACAAGGGAA 240

TCTACGGGTG ATAAAGATAA GACGGTGAGA GAGATAAGGC TAGATTGTAT AAGGCTTGAC 300

AGACCATAGC AAGATAAGCA AGGACCTGTG TCCTGTTAAC CATTT 345

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATAAAATTGG TCTGGGTACC CTAAGGTGTT TGCKTTGATA GAAAATTGAC ACCCCAAACT 60

AAGTGTTCTA CTTAGCTTCT ACAATAGTTA TTCCTAGACC TTAGATTAGT CATTACATTT 120

TTATTTAAGG TACTATGTTA CTTTCATGAC TACAAAATGA GGCACTCGTA CAAAACAGGA 180

ATGAAAACAT ACATATACTG TCTTGTCTTT ATGTCGTATT AATGCCAAAG ATATTGTCAG 240

GGATTATTTT AAAGAAGCCC TTACTCATGA TGGCTATTTT TAAAAATGGC ACAGGACAGT 300

AACAGGCTGA AAAGAAACAC CTGGTTTGAG GGGCCAAATT AAG 343

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 153 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ACAGGATGGT CAGGACAAGC CACCTCTGGT AAAGTGACAT TTGAGANGAC CCCTGAAGGN 60

GGGGGGTTGA GTCATGTGGA CATCTTGAGG AAGAGTTTAC TGGCACAGGG AACTGCAAGG 120

NCAAAGTCCC CAAGTACTAG GGCTGGGGGC AGT 153

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 257 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TCAGTCAGCT TATCGCAGGT GCAGCCAAAC ACAAAGCTTC AGGACAAATT GTACAAACTT 60

TACAATGTGG GATTTAAATT TAAAATATGA TACATAAAAA TCTACACAAA ACTGATAAAA 120

ATCAAGCACA GNTACCAGGA TTGAAACTTA TAATAATCCA TGTGTGAAAG GGAGTCTTGT 180

TTGCTTTCAA GTGCTTTTAT TCTGCTATGG AACAGTCAAA ATGGAAGNTG TAAAGCTTTG 240

TGGTTAGTTT AAATTAT 257

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GTGCCACCCA TATCTAATCC AACAAGTCCA GCTGCCTCTC TCTNAAAMAAT ACCNARGATC 60

AGGCCCCTTC TCAGCACCCC CACAGCTGCT GCCCCAAAGG AAGCCACGTC ATCTCTCACC 120

GAGATTGTKC AGCAGCCACT GCCTCCTTGT CACCTTCCCC TGTGGTCATT CTCCCCACAT 180

GGCCAGGGAA TCCGTCCTGT TAAAGTCTGC TAGGTCACGG TCCTTCCTAC TCAAAATGCT 240

CCCYTGGCTC CCACTGCCCC CAGAGTAAAA AGCCCAGACC TTCAAATGAC ACAAAGCCCT 300

ACAACGA 307

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TCCACAGCTC ATCAGATGCC TCTTNGATAA TATATAAATA GTAAAAACAA CTTTTACTTT 60

TTCCTATTNT AATCGTGTGC CATGGATCTG ATCTGTACCA TGACCCTACA TAAGGCTGGA 120

TGGACCTCAG GCTGAGGGCC CAATGTATGT KTGGCTGTGG GTGTGGTTGG GAGTGTGTCT 180

GCKGAGTAAG AACACGNTTT TCAAGATTCT AAAGCTCAAT TMAAGTGGCA CATTAATRAT 240

AAACTCAGAT CTGNTCAAAA GTCCGG 266

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 388 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CAGCTTTGGA AAGACTTTGA CCTCTGAACA AAAAGCCAGA AGGCTGCTTA AAGAAATAGT 60

AAGGGTTTCA CTTGCCCTGG ATAGTCACAA ATCTAGGAGT ACTGGTTCAC TGCCTTGGGT 120

TACCAGGTAT CAGCTCTTTC ACAATCTCTC CTCTTCCCAT GCTTCCCCTT AAAGTCCAGT 180

TGACAAATGA AAAAGAAAAA AAGGCCTTGA TTTATAGTAT TGCCAAACAA CCTCATAAGA 240

ATGGGTAAAA TTACATACAC ACATACATAG AGAAGGGAGG TAATGCTGTG AATCTACTTG 300

AGCTGGATTG CATGCTCCCT AGGGACCACG GTGCCCAACC TGTAATTTTA TTTCTAACTT 360

TTATAAATAT ACTCCTTTTT CACGGATG 388

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GAATGTCTAC ACAAGGAAGT ACAGGATTTG GCTTTTCTAG ATGTCATATC CAAACTTCGC 60

AGTCATGAGA ACAAAAGTGT TGCCCAGCAG GCCTCTCTCA CAGAGCAGAG ACTTACTGTG 120

GAAAGCTGAG AACTGCCCGA TACACGGCAT CATCCCATCT CTAATTTCCC CTCTGTCCTC 180

CATCCAGCGG CTTCTTCCGC TTCATTCTCT ACCATACCAC TTGTGCATGC ATGTRATGTT 240

CTAATACCAA TTGAAGAACC GCTGTAGGTA CCTCCCTAAT AAGGATTTCT AAACCTATAG 300

TTAGTGTGAT CATGACTTTG GTCAAAGGCA AGTYTCCCAC CC 342

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATGACTTGG AGAATGCCGA AGAGGAAGGC CAGGAGAATG TCGAGATCCT CCCCTCTGGG 60

GAGCGACCGC AGCCAACCAG AAGCGAATCA CCACACCATA CATGACCAAG TACGAGCGAG 120

CCCGCGTGCT GGGCACCCGA GCGCTCCAGA TTGCGATGTG TGCCCCTGTG ATGGTGGAGC 180

TGGAGGGGGA GACAGATCCT CTGCTCATTG CCATGAAGGA ACTCAAGGCC CGAAAGATCC 240

CCATCATCAT TCGCCGTTAC CTGCCAGATG GGAGCTATGA AGACTGGGGG GGTKGACGAG 300

CTCATCATCA CCGACTTGAG CTGGAGTCAT CTTTCCTGMC CTTTGCCCCA TGCCC 355

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCCAAAAAACA NYTCTGAACC CGTTTTGGGA AATAATGGGA TTCCTTGATC ACGGGACAAC 60

GAATCACCCT GAAGTTTTTC TCCAGTTTAC TCAGTCACAT AAGCCACCAG AGGCTAACCA 120

CACTGACAAC AAAAGCAAGT CCCAGGATTC CGGGGGCTAA TACCATGCTA GGCATTACTT 180

GGCAAGTTAT GACTTGGTAT ACATCTGTGA ATTTGGTGGG AGGAGAAAAC TAACAGTAAA 240

TTTATCAAAG CCACTGGTAC GTTCAGCGTT ATAAAAATTA CAAGGATCTG CTTCTCGGCG 300

ACT 303

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 178 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGTCTGGGGG GCTAGAGATA CACATGCCAG TNCTATACAT TTCTCAGCAC TGTGCTGTCC 60

ATTCACAGCA GTTCAATTGT TCATGCGATA TAAGCCAGTC ATCTGGCCCA AGTTATTCTC 120

TGGGCTGTGT TCTCTGCAGG AATCTGATGC AAGAAGGCCT GAAGGATGCA TGCCTTTT 178

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 322 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TGCCTTTCTT TAGAAATTTA GGGCAGTGTG ATGCTTCCAG AGGTCTGTAC AAACACCAGC 60

TTTCATTGTG CTTGGGAGTT TCCATGCCTC TYCCTTCTCT TCGCTTAGTG CACGTTTCTG 120

CTTTTTATCA GTTTGACTGC CTGAGACTGA KTCCAACAAC CCAAACTGAA CGCTCAGCTC 180

CTCCKTTTCA AAGGAGGATG ACTTNTCTNA ACAACTATTT AGGTGAATTA TTKCKACAGT 240

TTATTAAAGC AATGGCTCTA AACAAATTCC ACTGGGGGTG ACAAAGTACA ATACAAAAGG 300

CGTACTCTGA GGGCTTGGGG GT 322

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AAACTTTGGC ATTTTTATTC AGACACGTAT AAAAACAAAA CAAAAAACTT CAGTGATACA 60

ACAGACGTTT TCCCTTAGTT CCCCATCCAA GGGGACAGAG GTGTGCAGCT GAAGCTGGAY 120

CTTTTTTCTG TCCTACCTGG AAGCTGTCTC ACTGCTGGAT GAGAATGGCT TCTAAAAGTG 180

GATCTTGGGG ATCCTTGTGA ATTTGCCCTC GGATAAGGAG TGAAGWTCAT TTACGGCACA 240

TGTGGATTAT GGTTTACACA AAGATGTCCA GTTATTTT 278

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AGATCAAAAG ATGAGAGAAG CTGAAACAGA ACCGCATGAG GGAAAGAGGA AAGTGGAATC 60

TCTGTGGCCC ATCTTCAGGA TCCACCACCA GAAAACCCGT TACATCTTCG CCTCTTTTAC 120

AAGCGGAAAG CCAGCAGCAG GATCTCTAGG AATATTAGTA TTAAAGAAGG CTATGCAGCA 180

TAAACCTGAT TTCAAAATGG TAAAAGCAAG GTTATGTGTA CTTGT 225

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGATTGCCAG GAGCTGTTCC AGGTTGGGGA GAGCCAGACT GGACTATTTG AAATCCAGCC 60

TCAGGGGTCT CGGCCATTTT TGGTGAACTG CAAGATGACC TCAGATGGAG GCTGGACAGT 120

AATTCAGAGG CGCCACGATG GCTCAGTGGA CTTCAACCGG CCCTKGGTAG CCTACAAGGC 180

GGTGGTTTTG GGGGATCCCC ACGGCGAGTT CTGGCTTGGG TCTTGGAGAA AGGKGCATAG 240

CATCACGGGG GGACCGGAAC AGCCGMCTGG CCGTGCAAMC TGCGGGGACT GGGATGGGCA 300

AACGC 305

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 264 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATGAAATAGC ATATCTNNGC CTAATTAAAA GATTCCATTA CATTTACTTT TATCATTTAT 60

ACTGCCAAGG ATCAGTCACA AAAAATTCAA ATTATACATA TTATTCATGC TTTAATTTCA 120

TAAATAAGTA AATTAAAGCA AGCCAATATG TCTCTCTTCA TAACATAGGG AAAAATTACT 180

GTTTAGCATA ACAGNGTAAT AGGCAAAGTC TAGCCATACA GCAGCAGTTC ACGGTGTTGT 240

CAAGTTGGKA CAGGTTCCAT CGAT 264

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GATCTCTTCC AGCGTCAATG TACTGGGACA GCAAACACTC ACATTTGAAG TTCCTTCTGG 60

CCACCGGCTT CCCAGTACAT TGACGCTGGA ACAGATCATC TCAAATGGTT CTCCAGTGTC 120

AGCCTGGAGA TCTCAGAAA TGGAGTCTAC TCCTGGGGTG GCTTGTATGG GAGCC 175

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTCTCTCAGA GCNACCTGGGC AGCTCAMRCC CACAGCGGCT CCTCATCCTC TGTCCTCCCA 60

TCCTCATTCC ACTCTCATCT CCCACCTYCT CACCCTGGCC TCTAGCTTTC TCATCTACTC 120

TAGCAATTCC TGTTTCTCCT GCTGTAACTG CTCCTTTTCC TTCTGGAGCA CACGCAGGGC 180

TGACCGCAGC TGTGTCAGCT TCCGCTTACT TTMTGACAAC TGTACCAGGC TAGAATCCTT 240

TCTGCCTGGG TCAGCTTCAG TCTTTGAACA 270

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCCTGAAGAG TGGGTGGGAC AACCAGATGG GTGTAACCCC TTGTGGGGGA AAAGGAGTGA 60

GTTTACTTGG TAAAATAATA ATGGTAATGT CAGCAGCGTG GCTGGGGGAC TCAGTATGGT 120

CCCGGGAAAA GAGTTGGGGC AGTGAACTTC CCAGGCCGAC TGGCCTTGGG CTGGCAGCAG 180

GGAGGCTGCA GGGCGCCTAC CTMCTCTGCC ACGTCCCTGC CTAGGAAACC TATCCCAGGA 240

CACCCTGCTT TGGCCTGGAT AGCAGCCTAG GGATGAGCAT TTCTTTGAAA GCAATTAGGT 300

TATTCACCTG GTATTAAAAC TATTTACTGT TAAAAAATCT GTGACTTCAT GGARGTGGG 359

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCAGGAAGGA CAGGAAGTGT CCTCTAATAC GCATAAGATC CAGTACAGGA GAGATGGGAA 60

GMGAGKCTCC AGGATGAAGG GGAAAARAGG CCGCATGCCA GTCACCTGGC ATCTNCCAGA 120

GAGGGYCAGY CTNCCCACTG AGACTGGGGC ACGAGTCCCG TCATCACCAT GCCCTCTGAC 180

TGTCGAACTG TCTTTTTACC TGACAAATAC TACACAGGTA TCGMTCGTGG CCATACTCTG 240

CTATCTAAAC CCAGGAACTG ATTAGATTGT T 271

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CTCCAAGCAG TAAAGACTTG CAAAGCATTG CATTTTGATT AAACCTTGCT GGGCTGAAGG 60

GCAGGCAGAG CTGTGGTGGA CACTGGCAGG ACGCAGCACC CCCCGACTGG CCCTTGGCAG 120

GCTGCACCGG GCGCATGCGG GTGTGGGCCA GGGTTGCTTT AGGAAGCAGG TGGGAGTCTK 180

NCACGTGCAG KCGGTCCAGG AGKGYACCAK GCCTGGCAGG GCACTG 226

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 408 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGTGGGGCAA GGTGGGGGTG AAGTGCACTC CTGCTGCATG AGTGGCAGGG CAGGGTGCAC 60

ACACACACGT GGGTMCTGGC TGGGTGAGGC AAGCAAAACC TGCCTGCACA TGGCAAAGGG 120

ATGTGGGAAG TATCCATGGG CNCCAGGGGA AGCTGCAGTT TGGGGAGGGA ATGGGTGGCA 180

CTGCTGCGTG TCTGTGGGGG CCACCCCACT GGGGGTCTCC AAGTGGTCAA GTTCCGTCTG 240

CCAGGTTAGA ACCTATGATG GGGGCTTCTA GGACACTNGA GGCTGACCTG AAAGCAAGGT 300

ACTTTTCACA CTGGGACCCT GCAAGAGGCC AACAAGATTA AGGGATGCTT CAGGTCAGAC 360

TTGGCCCTCT TCTTATGGGG CAAGACCTTC CCCGCAGAGT TCAGATCT 408

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TTCTGTGCAG GAGGACCACA TGGCAGTCCA GCAGACTGCA CATTTTTAAA AACTAGGTCT 60

TCCCAGGTAG TTTGAGGAGC ACCAGGGCAC ACTCAGGGAA GGGACATGTC AGTGTCTGAG 120

AGCTCACGGG AGGAAGGTGT AGTGACAACA TGGACCATGG TGGAGTGACT TTAGACGGCT 180

CTTGGGTNAG GAGAATCATC ATGTAACAAA GCATTAAATC ATTTGGAGAA ATTCAGAAAA 240

NTCGTAGATG TACATTCTAG CCCACTTACC AGGCCTACTA AACGTCAATC AGATATATTT 300

CAATTTGAAT TCGG 314

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

AACCCACCCC [illegible] TTACATTTAT AATCTTATAA [illegible] GGCTCCTCCA 60

CTGGAGCAGT GGTTCTCAAA CTCGTGTATG CATAGGAATT ACCTGAAGGG CTTGTTAAAA 120

CACAAACTGC AGGGCCCACC CCCAGAGTTT CTGGTTGGGG AGGTGTGGGC TGGGCTTGAG 180

GATGTGAATC TCTCACAAGC TCCCAGGTGA GGCTGCTGGT CTGTGGACCC ACTTCAAAGA 240

CCCAGTGAAT CAGAAGAGTC AGTGAGACTG GACAAATGAA CGCAAGACAG TCTTCAAAGG 300

AGACCAGAGG 310

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TTTTTTTTTT TYCCGGGGAR GTCAAACATA CTTTTTCAAC ATAGGATKTC TGACAGGAGG 60

CCCTTGGMCA GGGTTCCCTG ACCTCTGYTT CAAACCCCAC TGGAAACAGA GCAAAGTCAT 120

CAMGAAAACC CAGGACACCA GGGCAGGGGG GCTGCACAAG GTCGGGTAGG TCACAGTGGG 180

CCAGCACACA GTGGCCCCGC CCAGGTCCAG CCCAGCCTGG GGGAGGGTGT GAGGGTTCCA 240

KGCAAGCTCA TT 252

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GTCAAGTCTA CCATCATTCT AGAAGGAAAA GGCATGGTGG GAATTCAGCA CCTGAACTTG 60

TATTTACACC AGCCTCGGCA TCTGGCAAGG RAATAGCGAT TGTTCATAGT GATGCAGAGA 120

GAGAACAGGA GGAKGAAGAA CAAATACACA CAAACAACTG ATCTAGGGAG ACTCCAARGA 180

TCCAACAG 188

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 304 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

AATCAGCCTG CAAGCAAAAG ATAGGAATAT TCACCTACAG TGGGCACCTC CTTGAAGAAG 60

CTGATAGCTT TTACACAGTA TTAGATTGAA ATAATGGACA GAAACACATT CTTGTCAAGA 120

AAGGGGGAGA GAAGTCTGTT TGCAAGTTTC AAAGCAAAAA GCAAAAGTGA AATGATTTGA 180

GGATTTCTGT TCTAATTGGA GATGATTCTC TGGTTGTTAC AAATGGCAAA TATTGATGAT 240

TGTGTGCTAT TGATTGGTGC AGGATACTTG GTATACGAGT AAATACTTGA GACTCGTGTC 300

ACTT 304

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CCAGAAGCTT CTTGCCTTCT CTGTGCTCTC AGTGGTTCCC TTCCCTGAAG TGCCTCCCTT 60

CTCATTAATT ATAGCCTGTG TCTGAACATT GTGAGCTATA AGAACCCTCA TATTAATGGT 120

TAAGGGACTG TTGGAAATGA TGTGATTTTA TTAAAAATGG GGTCTTTGTG GAGGAGTCAG 180

GAATGGTCAA AATGAGCTTC AGGTATGGGG CTTGCTCTRT GCTCCTGATA CCAAGGGTCT 240

GGCAAGCACA AAGGAAGGTG G 261

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

AATACGTATT CTGAAGCCAC TATATCTGCA TATGTATCCC AGATTTGAAC AATTAAGTAA 60

AAAGATGGTG AATGATGAAA GCCAGTTTTC TGTCTGTAGA AGTGAGAGGT GACAGATAAC 120

GAAAGGAAGA AGGCTAGAAT GGATAGAGGA CAGTGCTTAA GTGTAGTTCC TGTTGCCTTT 180

AGTCTTATAC ACTTCATTTC CAAAGTTTCT TAGCACCCCC CTTCCCCCTT TGGTGAGGTT 240

GTTTGACATA TTTTCTAGAC AATTAGATTC TTTTGTCAAA GTCTGTGTTC CATCCCGAGA 300

GCCTCTGATC TCTTAAATGA TTTTTTAAAT TTACATACAT TAAGGTTCAC TCTGCTGTAA 360

AGGTCTGTGC GTTTTAATCC TGTCTCACAC TTTTTGCATA TGTTCCCCTT ITGCCTGGGA 420

ATACTCTCCC AGATATTCCC CATGACTCCC CCCTTATCTT CAATCACATC 470

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 466 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GTGTTTCAAG GGAAGGCAAC TMCAAGTTTG TGCAGCTGAA TTTCTGTAAA GTTAAGACAG 60

ACTCAMCTTC TCATTCAATC TGGGGCAGTG GATAACCTTT CTGAATAGAC CCACTTGTTC 120

ACGGACAGGG ATAGAGGTTT GCCTTTCTTC TTTCCTTGAA TTTGGAGTGA GCACTAGGGA 180

GGGGAAGTGC ATGGGTGACA TGAAGAAGGT GAAGATGTAG TAAAAGCATC ATCCAGGTAC 240

ACATTAACGG TGCTGCAGAA TTTTCACAAT ACAACTGAGG GAGTCTGTAG TGGCAAAAGC 300

AATTACTGAG CACAAAAGCC AGTCCTCAAG GGCTGATTCC ACCTTCCCTG TCCAGGGACT 360

TTCTCAGCAA ACTTTGTTCA TGAGCAGTTG TTCGCTTTGA TGGTCTTAGC CAGTTTTTGG 420

TGCAGGGGTG TTCCTCTGGT ACTAGGGCTA GGGCAGCTGT TTAAAG 466

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 491 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GACACCCCTC CTGCCATGAA GAATGCCACT AGCTCTAAGC AGCTCCCACT GGAACCAGAG 60

AGCCCCTCAG GGCAGGTCGG GCCTAGGCCA GCCCCCCCGC AGGAAGAGTC CCCTTCCTCT 120

GAAGCAAAGA GCAGAGGACC CACCCCACCA GCCATGGGCC CACGGGATGC CAGACCTCCT 180

CGAAGGAGCA GCCAGCCATC TCCAACAGCA GTGCCAGCCT CCGACAGCCC TCCCACCAAG 240

CAAGAGGTGA AGAAGGCAGG AGAGAGACAC AAGCTGGCAA AGGAGCGGCG AGAAGAGCGT 300

GCCAAGTACC TGGCGGCCAA GGAAGGCAGT GTGGCTGGGA AGGAGGAGAA AGGCCAAGGT 360

GCTGCGGGAG GAAGCAAGCT CCATGGAGCG CCGCTGCCGG TTTTAGGGAG CAAACGTCTT 420

AAAGCCGAGC AACGCCGTTC AAGCCTTGGA GGAACGGCTA GCGGAAGAAG TTTGTGGAAA 480

ACAAGGGGCG T 491

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

ATCATTGAGT ACGCAGAGCT CAAAACAGAC GTGTTCCAGA GCCTGAGGGA AGTGGGCAAT 60

GCATCCTCTT CTGCCTCCTC ATAGAGCAAG CTCTGTCTCA GGAGGAGGTC TGCGATTTGC 120

TCCATGCCGA CCCTTCCAAA ACATCTTGCC TAGAGTCTAC ATCAAAGAGG GGGAGCGCCT 180

GGAGGTCCGG ATGAAACGTC TGGAAGCCAA GTATGCCCCG CTCCACCTGG TCCCTCTGAT 240

CGAGCGGCTG GGGACCCTCA GCAAATCGCC ATTGCTCGCG AGGGTGACCT CCTGACCAAG 300

GAGCGGCTGT CTGTGGCTGT CCATGTTCGA GGTCATCCTG ACCCGATTCG GAGCTACCTT 360

CAGGACCCAT CTGGCGGGGC CACCGCCACC AATGCGTATG ACGTCGATGA GTTTTTGAGT 420

TCACTGCTGT GAGCGCATGA GTCGTGTACT GAATCCTGTG GACAACGGTT AAGTTACA 478

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CCTGGAAAGT GGGGGTGGGC CAGGGGGCCA GGCCCAGCAT GCACCCCCAT TTTTTTGGGG 60

GGTGATCCCT GCCCCAGCTC TGCTGATACC CGGGGCCACA GCGTCAGGCC GTTGGGGGTG 120

GAGKTAGAGG TGGGAGAGCA GGGGAGAGAG CCTKAGGAGC CACAATTGGG CAGACAGAAG 180

CGG 183

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGATATTGCA CCTTACAGAC TTAGGGAGCC TTTACCAGAG ACGCCTAAAA GGCCCCAGGT 60

TCAGCCATTG TGCTGAATAG AGTGGAATAT AGAACCAGGG ACAGAGTATT TCATTTAACG 120

TTGATATATA CTTGCTAAGG AAACACTAAC AATACTGTAA CTTTGTTAAA GGACATAGTA 180

TTGAAATGGG AAATAGAGGT CAGGCTCACA TCATCTTAGT TTAATGCTGG GCAACTTTTT 240

CTGATTTCTG TAGTTCCCTG GAAAATGTGT CCTTCGTACC CATAAAGTGG TACAAATGCA 300

TTTGTAACCA TTTTTG 316

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

ATCTGGTCTA GAGAGGCGAC TCCAAGCTCT CTTGCTGGCT CCCAGCTGTG GGAATCCTTT 60

AGGCTTGTTC TCAACCTACA CGTTAAAAAT GCTTCTTGGT GTGTTTGGGG AGGGGGAGAG 120

GGAAACTGAG CTCTCTCTTG ACCTCCTCCA ACACCCTTGA CTTGCTTACC CAGCCATTTT 180

CAGTAGCTAC ACGGGTGGTC ACAGAACACT GGGCGGCACT CGGCACACAA CACAGAACCG 240

GGGCAGTCCA TGCAGGTGCG GGAACACATG TCGGACCCAG GGAGCAAGGA ACACGCCACC 300

CCGAGGAACA TGCAAACGGA GGAAGGATTC CCTTCAGATT CCAAGGATGC CACAACCCCG 360

ACGGGCGGCT TAGGGAGGCA CCGATTATCT AAGGAAAAAG GCCACTGTTT G 411

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 413 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CTGCTCCTTA TGTTTTTATT TCCAAAGTTT AGAATTTCTT TGCTTCATAG TATTATTTTA 60

TTTTACTAAA TTACAGAGTA AGAAAAGCTT TTCATTTTAT CTGATTTTAT TCTTAGAACA 120

AAAATATTAC GATCTTCTAT ATTTTTGTTC TTTTGCCAAA AAGTGTAGGC AATTTTACAT 180

CATCTTTTTT CCCAATCAGT TTGTGATCCA ACTATAAAAA GGAGACATAG AATACTGAAT 240

AAATGAAACA GAAACTCCAA GGCCAAGAAG TGTCCATCTT GAAAGAGTGT TAGTGGCAAG 300

ATATGTGACT GCAGACTAGA TGTAGACAAA CCTGAGAAAA ACCAAGCATG GGGGAAAGGA 360

TYCCTATTTT AATAAATGGT GCTGGGGAAA ACTGGCTAGC CATATGTACT TTA 413

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCACGGTTAA AAGACCAACG TGTGTGGNTC AAATATAAAG GCCACACCTT TCAGACCGAA 60

CCTACTCAAA GATCCTTTAC TTTGCAATAA TTTGAACTGG AGAACCAAAG ACGGGAGACG 120

AATGAAAGCA AAGATGCTCA AAGAACCAAA GGAAAGACCT GAAGGAATCC ACCTGCATAG 180

GCCACGCGTT CCACTCTGGG TCAAATGCTT CCACGATGCA GAAACCTTTT TTTAAAAAAG 240

TGCAAGTCTA ATTACCTACC AAGGGTAATA AAAAGCACAG CACAGGAATG ATTACAGCTG 300

ATGGTCAAAA AACAAACCAA AACCATTAAA AAAACAATCA GGCAGAAAAC AGGAGTTAAA 360

TGTTTACATA TG 372

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 389 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TCTAGAACCT GGACCCACCC AGCGCGTCCT TTCTTATCCC CGAGTGGATG GATGGATGGA 60

TGGATGGTAG GGATGTTAAT AATTTTAGTG GAACAAAGCC TGTGAAATGA TTGTACATAG 120

TGTTAATTTA TTGTAACGAA TGGCTAGTTT TTATTCTCGT CAAGGCACAA AACCAGTTCA 180

TGCTTAACCN TTTTTTCCTT TCCTTTCTTT GCTTTTCTTT CTCTCCTCTC ATACTTTCTC 240

TTCTCTCTCT TTTAATTTTC TTGTGAGATA ATATTCTAAG AGGCTCTAGA AACATGAAAT 300

ACTCAGTAGT GGATGGGTTT CCCACTTCTC CTCAATCCGT TGCATGAAAT AATTACTATG 360

GTGCCCTAAT GCACACAAAT AGCTAAGGG 389

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 329 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GAAAAAAATGG GAGGGCAGCC ATGTATTAAT TGTACATCCA AGGAAACTGT GCCCCAGGGG 60

TGTTGTGTGT ATTTCTGAGA AGAGGGGTGA GAAAAGGCAC TGTGTCAACA TTTGCTTCTG 120

CCTGAACGTG CACCTCCCAG TGCTCCTCCA TCAATTAGGA GAACTGTCTT GAAGAATGCT 180

GCCTCAGCTT CTGAAGAGAA GACCCCAGGA CATGCATTAA TGAGAGGAGG GGAGTCACAG 240

CTGCAGAAGA ATAAAGCTCT CTGAGGGAGC CTGGGNGCCC CCAGTGGAGG CCTGGAGCTT 300

GTTGACCANN GCAGCAGGAG ACCCCTGCT 329

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TTGAGTTGCC TGAGGTCATT CACATGCTTC ACTATCACTT TTCATCTGTT CAGCCAAATG 60

CAGTGCCCTA CCTCCAGCAC CTGTCCTTTG GTTATGATAA CTTAAGCATG CAGCTTTGTA 120

CCTTACCCCC AATCAAGCAT CTCCTTCACC TTTT [illegible] TCAACTTTTC CAACTTCAGA 180

AGAATGCTTG TGGTGCCCTT CGAAACCTCG TTTTTGGCAA GTCTACAGAT GAAAATAAAA 240

TAGCAATGAA GAATGTTGGT GGGGATACCT GCCTTGTTGC GGCTGTTGAG AAAAATCTAT 300

TTGATGCAGA AGTAAGGGAG CTTGTTACAG GAGTCTTTGG AATTATCCCT CATGTGATGC 360

CTGTAAAAAT GACATTCATT CGAGATGCTC TCTCAACCTT AACAAACACT GTGATTGT 418

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 336 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CTGAATTTTT ATATGCTTCA CTTAGGCTTT CATTTGAGTA GACTCTAAAA ATTCTGCCTT 60

GCTTAAGTNC TAACACTGCC TCTCAGATTT CAGTTTTGGA CATTGCACAA CTAAGACCTT 120

TTAAACGCAT TTNCTTGCTA ACTCGGAAGA CACATAGTCT GCAGCAAGAC ATTCCTATAT 180

TGAAGAAATG AGAGAAAATT TTATGCTGCA TCAGGTGGAG AGCAAGGCTC AACGGTGGTT 240

GCATTAGTTC CCTCGGAAGT ATTGAAAAAN CTTTGAAATG GGAAGGAAAA TTTTTTGCAC 300

CTAATGTTCC TGAGGTACCC AGAATGTCTG GGGGTT 336

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTGCTCAGTA AATACAAATT GGATGGACTA GAGAGATAGC CCCGAGGACA CTGCCAAATA 60

AATAACAAAT TGTGCAAGCA GCAGGCCGCT GTAATTAGAC CAAGGAGGAC AGTCAGTTAT 120

TAATATCAGA CACGTGGCAG GGTTAACAGC CACTGAGGGT GGGTACAATG AAGAGAGTCA 180

CTTTCTGCAC CCTCAGGGAC TTCGCTTGTG ATGGCCTTCT AAAGAGGGCT GAACAGCACC 240

AAGTGCCCTC GCTGCCTCTG GTTCCTGCTG CCCTCCGCGT GCCTTGGGTG CCCCACAACT 300

AGGGCCCTGG GTCCCTCCCA TGTCCCCCTC CCTCCTACAA CCCCTCAGCC CCTTATCTGG 360

CCAGCCATTA TGATGCCTAT CAGTATGAGG CCAGATGAGA GT 402

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 454 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

```
GGACCCCGGG CCCGCGATGT GGCCCAGTAC CTGCTCTCAG ACAGCCTCTT CGTGTGGGTT    60
CTAGTAAATA CCGCTTGCTG TGTTTTGATG TTGGTGGCTA AGCTCATCCA GTGTATTGTG   120
TTTGGCCCTC TTCGAGTGAG TGAGAGACAG CATCTCAAAG ACANATTTTG GAATTTTATT   180
TTCTACAAGT TCATTTTCAT CTTTGGTGTG CTGAATGTCC AGACAGTGGA AGAGGTGGTC   240
ATGTGGTGCC TCTGGTTTGC CGGACTTGTC TTTCTGCACC TGATGGTTCA GCTCTGCAAG   300
GNTCGATTTG AATATCTTTC CTTCTCGNCC ACCACGGCGA TGAGCAGCCA CGGGTCGAGT   360
CCTGTCCCTG TTTGGTTGCC ATGCTGCTTT TCCTGCTGTG GACTTGCGGC CGTTTGCTCA   420
TTACCGGGTA CACCACGGAA TGCACACCTG GCTT                                454
```

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

```
GCTTTGATAG CTAGTTGTCT AAAAGTGCTG NTTATTAAAT AATCCACCTN TTTCCCCACT    60
TAAAACATCC CTCTTACCAT ATACTAAATT CCNGTAGCCC TGGGTCTGTT TCTGGACTCT   120
CCCGTCTGTC TGACCCCCTC CAGGTCACAC TGAGTGAGGT AATGGTGGCG TGAGAATCCT   180
CTGGGAATCT GGCAGGNTCA CCCCNGAGCA GTCCACCCCN CAACTCATTA NCATCGTTCA   240
GAGTGGNCTG ACTGNTCTCA CACATTCACT CTGCCAAATG CACTTTAGGA ACTGTCAAAT   300
TCCAAAGTTT CAA                                                       313
```

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```
CTCAGCCGTA GCCCTAAGTC GTTTTTCCAA TTTAGGAAGC TCACAAGGCA GATCTGCATT    60
GTCACGTACC AGCTGTTTGT GAACCTTTGT AAGCTGTTCC AGGTTGTTCT CAAGAAAGGA   120
AATCTTCTGC TTTTGGGAGT GAATCCCCCC ACTGTCTTCG GGCTCCATTT CTGCACTTTT   180
CTTGACTCGA GTCCTGACCT CTTGAACCAA CAGCTTCCGA AGGTTGTGGC GGGTCTGGAG   240
TTCCCCGGCA ACTGTCTCCT CCAGACCCTT GAGGTCCTGC TTGTGACTTC TCAATGTCCT   300
```

TCGTACAGAA ATGTCAGCTC CTGCAGCTTT GGTGCTCTTC TCGTGGTTCT TCGCTCTTTC 360

AGCTTTCTCG TAGTCAAGCC TGAAGGCTTC TCTAAGCTCT AACTGGAGCT TCTGATTTAA 420

GGTCTTTTGA GCTCATCAAA TGGTCT 446

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 296 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

AGCCGGTGGC GCAATGGAGA GAATGTGCCT GAGACAGAGC GCCTGGCTGG GGAGGAGGCA 60

GCCCTGGGNG CCGAGCTCTG TGAGGAGACC CCTGTGAATG ACAACTCATC CATCGTGGTG 120

CGCATCGCGC CCGAGGAGCG GCAGAAATAC GAGGAGGAGA TCCGCCGTCT CTATAAGCAG 180

CTTNACGACA AGGATGATGA AATCAACCAA CAAAGCCAAC TCATAGAGNA GCTCAAGCAG 240

CAAATNCTGG ACCAGGAAGA GCTGCTGGTG TNCACCCGAG GAGACAACGA GAAGGT 296

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CCTTTCCTGC CTGGGAAGTG ATGACTCGCA GGTCGGGCTT GCGGCTGGGG GCTCCAAGCT 60

GGGTGCTGTG GGTAGGTGGG GGCGGAGACT TGGCAGGGAT GACCTTGTTT AGGCTGTTGC 120

CATTGGCCAC AGGGAGGAGG CCAGGGGAAG CCCGAGCACT GACGTAGCCA TTCCCAACAG 180

GGCTGGGGCA GGCTCCGTTA GCACTGTTCA GGTCACCNCC CAGCATGGCC CCCGCACTAG 240

CTGGCCGCTG GGGCAGGCCA GGAGACACAC TGTTCCTCTG TAGTG 285

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

ATGATTTCTT GCCTGTNATA ACCTATGCAC TCACAAAGAT GAACTCTCTG AGAGGGATGA 60

GCAAGAGCTT CAGGAAATCC GAAAGTATTT CTCCTTTCCT GTATTCTTTT TCAAAGTGCC 120

GAAACTGGGC TCGGAGATAA TAGACTCCTC AACCAGGAGA ATGGAGAGCG AAAGATCACC 180

GCTTTATCGC CAGCTAATTG ACCTGGGCTA TCTGAGCAGC AGTCACTGGA ACTGTGGGGC 240

TCCTGGCCAG GGATACTAAA GCTCAGAGCA TGTTGGTGGA ACAGAGTGAA AAGCTGAGAC 300

ACTTGAGCAC ATTTTCTCAC CAGGTGTTAC AGACTCGCCT GGTNGATGCA GCCAAGGCCC 360

TGAAACCTGG TGCACTGCCA CTGCCTTGAC ATCTTTTATT AA 402

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 246 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CATTTTTAAT AGAGACGGGG TTTAACCATG TTGGCCAGGC TGGTCTTGAA CTCTTGATCT 60

CAGGTAATCC ACCCACTATG GCCTCCCAAA GTGCTGGGGT TACAGGTTTG AGCCTCTGTN 120

CCCGGCCCGG CCAAAGACTG CCTATTCTAA ACGTTGCTGA GGACGTGGAN CAATCACAGC 180

TCTCCTNTCT TTCCAGTGGG AGTTTAACAT GGCACAACCG CCTGAAAACC GTTTGGNGAT 240

TTCTGT 246

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GGGAACCCTC AGCAAAATAT AATGGTACCG CTATTATCAG CCTTGTTCGA GGCCCAGGGA 60

TTTTGGGGGA GGTCACAGTG TTCTGGAGGA TATTCCCTCC TTCCGTGGGG GAATTTGCTG 120

AAACATCAGG NAAACTGACA ATGCGAGACG AACAGTCTGC AGTCATTGTA GTAATACAGG 180

GTTTGAACGA TGACATTCCC GAGGAAAAAA GCTTCTATGA GTTTCAGCTC ACTGCAGTCA 240

GTNAGGGAGG AGTTCTGAGT GAATCCAGCA GCACTNCCAA CATCACGGTG GTGGCCAGCC 300

AGTCTCCCTA TGGCCGATTT GCCTTTTNAC ATGAGGCAAC TTCGAGTGTC AGAAGCACAG 360

AGGCNTAACA TCACAATCAT CCGTTCCAGT GGAG 394

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATAAGACCAT TGGCAAAGGG AGAATTCATG AACTGAAAGA TCTGAAGTAA TTTCCCAGAA 60

TGTAATGTTA AGAAATAAGT TAAAAGGCAG AGCATAATGA GTCTAACATG TGTGATTGAA 120

GTCTTATAAG GMGAGAATTA AGAMCAGGCA ATATTTTAAA GGRATAATGG AGAAAATGGA 180

ATAATTGATG AAATATGTGA ATATATATAG GGACCATATG CATATGAMGG CCGGGGGTTA 240

AATAAAACGA AATCTACTTG TACATACTTT ATGGGATTCC TGCAGCCCGG GGGGATCCAC 300

TAGTTCTT 308

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

CTTTAACTTA ATGGCAATTA AAACTCACTG GCAAAAAAAA TCACTAGAGA TGTCAGTCCA 60

TTATCTTACC AAATAGTGTA TTTTTACCAT CTTTTACCTA CACCCTTGAG TAAGGTGGAA 120

TAGGTTAAAG TTACTGGCAT AATAACACTT CATTGAATTC ATGATAGTAT TTAACATGTT 180

AAAACTGTTT AGTTGAAAAG TTCACATGCA ATTTATAATT TAAAAATATG CTACATATAT 240

TTCATAAAAW TACAATAGGT CATACTARAC TTTGACTAAA ATTAAGAATG TKTTTCTKTC 300

ATAATAATGC AGG 313

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TGCTCCGTTT ATTGCTCTAT TCAATGACCA CGAGCGAATT ATAAAAAGAC ACCAAATGTC 60

TCTGTCTGCC GTGGGATAAA TATTTAAAGT CAGCAATAAA GTCACGTGGC TCCAAGRTAA 120

TACATGTTGC CAAAGAGTCA TGCATGCCCT CCTGATGGGC TCTCAACACA CGTATGGWCA 180

TGGGAACACA CGCAGAGCAA CACGCAGTAT GAACTTSTGG GAAGGCTTTA CCACAGTGAC 240

ACAGTAAAAT GTCTCACGTA GATCTGRGCT GAGTCCCCAC CCAAACCTTG AGCTCCCCTT 300

CCA 303

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 380 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AAAACAAACC AGCTTTAATA CCAATATAGT TCTCTCTTAA ATACCGTGTT TCCCAGGACA 60

AATGCAGGGG CAGGCTCTTG GCAGAAAGAG TAGAAAGGAA ATGTGGAACA AAATGGAATG 120

GATGGCCCAG GCCCAGGGTC CCTGCCTTGG GCACTAGGGA CTGGGCTGCC TCGGGGATGG 180

GGGAGTGACA GCAGCTCCCC CTGGTCCAGT TATTGCAGAG GCGTCGGGGG CTCCCCTCCC 240

TCCCCAGGCC TGAAACATTT CTCAGGATTA CTTCTGACCT TCAGCCCCAG CAGGGCCAGG 300

GCCTGGGCTC CTCTGGTCTA GGATGGGCCC CTTTGCCCAA AAGGGCCTTC AGCTAAGGCG 360

TTGGGTTGGG CGGGGAGCCC 380

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 280 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GCCTTTGCTG CTTATTCGCA TCGATGGTGA AAGAGATGTC AGGAGCACTT CTCTGCTGAG 60

GTGGCTGAGA CGAAGAGGAC TCTGCTGCCA GCCTTGCCGC ATACCTGGCA ATTAGCCTGT 120

GTTCTTCATC AAGCCGGTTT GAACTCTCAA GCATGCTCCT GGTAATAAAA GGACTTCCTG 180

AGGAGGGAAC AGAGTGNGAG AACAGGGTGT CGTTCATGCT GGTTACAGGT CTGGGAGGCA 240

CGATGTGAGC CAAGTTGAGT GGCTTCTCAG GCTGATCTGG 280

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

CCTGCGTCTC TTACACCCNC TCCCACCCGA GGCTCCCCAG AGATAGCAGA GAATTCGAAG 60

AGGTCCCCGG GGACTGGAAA GAAGTCCCNG NACGCCGCCT TCCCAGTCTA TACCCCAGCC 120

TCCTTCCCAC CCTACATCCA GACCCAGCTC AGACCTTCGT GACCACCCCA TCCCTTTCTC 180

CGCCTGCCTC CGTCCCCCGC ATCCCTCTCT CTCCCTGCCT TCCAGAGCCA GGACAGGCCT 240

CCTCCTAAGC CCCCAAAGTT CCTCACCTCC TGACTTCCTC TCCCTTTTAT TAATATCTGT 300

ATTCCTGATA ACCCTCCTCT TGACTATCTC TCTCACCCTA TGTCCCAGCT TATCCACAAG 360

CCCGTGCCAC AGTGACCCTT CCCATACTTC TGGGGGGGCT GCTCTCCATC TGGATCGTAG 420

GAGGATATAG GTGTGTTCTG GACCAT 446

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

GTCCCTTCTG GGGACTCTRT TTCCCCATTT ATTGCTGCTG TGTCCCTNAC CAGTTCCTTG 60

CAGGATTCCC TCCTTTTAAA ATGCCCTTAA ATCTAGCTTT GCCTTGGAGA CCCCAGTGGG 120

TGCTGCTCCT GCCGTTTTCT TCCTGCCAAG CCTGAATCAA TGTTTCATCT CCAACCCTCT 180

GCCAGTTTGG CCCCTCAAAG CTTGGTGGCT CAAGACTGTW AGCCTGGCAG AGCCGCGNGG 240

TGAAGGGAGA AGCTCTTGGA GCAGGCAGGA TGCCACCGCT GCTTCAGCTT GCCTCCTCGC 300

CCAGCTACCC TTTGGCCCCA TTGGGCCCTC GTMTGCCTCT CCAGGATTGT ATGTTTCAAG 360

NCTTGTCCTG TGTTCCTTTG TCTG 384

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TCAAGCTGGA AAGGGCTACT ACCTCATGCT GGAAAGGGCT ACTACCTCAA GCTGGAAAGG 60

GCTACTACCT CAAGCTGGAA AGGGCTACTA CCTCAAGCTG GAAAGGGCTA CTACCTCAAG 120

CTGGAAAGAG CTACTACCTC AAGCTGGAAA GGGCTACTAC CTCATGCTGG AAAGGGCTAC 180

TACCTCAAGC TGGAAAGAGC TACTACCTCA AGCTGGAAAG GGCTACTACC TCAAGCTGGA 240

AAGGGCTACT ACCTCAAGCT GGAAAGAGCT ACTACCTCCA AGCTGGAAAG GGCTACTACC 300

TCATGCTGGG AAAGGGCTAC TACCTCAAGC TGGACAGGGC TACT 344

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GCCCCAGGGT GAGGGCTATG AGGGGTCAGG GGTCAGGTTC CCCAGGACCC TAGTCCTTGT 60

CCCCTTCCCT GGTGCTAAAT AAAAGTGAAT AAATACTAAA TAAATACAAC TGGGGCCCAG 120

GCCCTCCCTG CCTTCCCCCT CCCTCCTGTG ACCCGCAGCA GAGGGGGCAG TTTAGATGGA 180

GGGCTGTCTG TCAGCCCCTT CCATCCACTA ACCCATCACT GCCTCCCAGG GCAGGAAACC 240

AGGGCAGGGC CAGCCTGCGC ATTAGGGCAG AGAGGAGGGG CAGGTCTCAC GCCCACAGCC 300

CCTTCCCACT TGAGTCTTAG CATGAGGCAG CAACAGAAGC TCTCTCTTCC TCCCAGCTAA 360

GTCC 364

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

ATTTAATAGA AAATTAAAAT AATAAATAAT ATGAAACAGA CTGATAACGC TGAGCTGGGC 60

AGGCCCAGGC CAGTCTAGTA CAAAGTTAAG GAGGTAGGGA GGATGGTGGG GAGGAGGGGG 120

CGGACTACCC TGCAGGACGC GGGAGGCTGC TCAGACTGTG GTGATGTCAG GAAGGGCCGC 180

ACACTTTGGC ATGGACGATG CACTAAAAAA AGAGAAAG 218

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

GCTTTCAAGG GAACAAAGAA TCGGCCTGGC AGTGCCCTGG AGAAGGAGGT GGAGAGCATG 60

GGGGCCCATC TTAATGCCTA CAGNACCCGG GAGCACACAG CTTACTACAT CAAGCCGCTG 120

TCCAAGGATC TGCCGAAAGC TGTGGAGCTC CTGGGTGACA TTGTGCAGAA CTGTAGTCTG 180

GAAGACTCAC AGATTGAGAA GGAACGTGAT GTGATCCTGC GGCAGATGCA GGAGAATGAT 240

GCATCTATGC GAGATGTGGT CTTTAACTAC CTGCATGCCA CAGCATTCCA GGGGCACAGC 300

TCTAGCCCAG GCTTTGGAGG GCCCAGTGA GAATGTCAGG AAGCTGTCTG GTGCAGACTT 360

GACC 364

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 410 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

CTTCATACTA GAACTGTCTG CCATCTTTAT TTCTTTGTTT TCAGGAAAAT TGGAGAGAAA 60
AGTATTTCTT TTTTAAAAAT GATTATTATA CTTTAAGTTC TGGGATACAT GTGCAGAACG 120
TGCACGTTTG TTACATAAGT ATACACGTGC CATGGTGGTT TGCTGCACCC ATCAACCCGT 180
CATCTACATT AGGTATTTCT CCTAATGCTA TCCCTCCCCT AGCCCCCCAC CCTCCAACAG 240
GCTCCAGTGT GTGATGTTCC CCTCCCTGTG TCCATGTGTT CTCATTGTTC AACTCCCACT 300
TATGAGTGAG GGACATGCAG TGTTTGATTT TCTGTTCCTG TGTTACTTTG CTGAGAATGA 360
TGGCTTCCAG ATTCATCCAT GTCCTTGCAA AGGCATGAAC TCATCCTTTT TATGGCTGCA 420
TAG 423

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

AACAGCCCCC GATCTGCATA GCCTGTGAAA GCCCACGGGG ACATCAGTAA CCTTCTGCAG 60
CCACCATCCA ATGCCATTAC TGTNAAGTGA GACTTGGCCA CTGTAGCCTG GGCCTGCTGC 120
AGGAGCTCTT CAGAAAGGCA CATGAGGACC ACGGTTTGCC TCAGTTTCTG GTAAAACACA 180
AGGTCTGGAG TGCCCCTGCA AAGGGTATTG ATGGACTTCC TGCCAGTGAC AGAGCATGTC 240
TATTGCAAAC AATTCTCTCA GTTACGTTCA GCACTTAAGA ACGGCTAATG NCAATAGGAT 300
CTTTAGCAAC TTTTTCACAT CATAGAAGGT GCAATCGCTC ACTTGGGAAC ACTACTGAGA 360
GTGACTTCTC TTTTAAAATT GAGTAGCAGA TGAAAAATTA AAATT 405

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

GAAGACAATA CTGATGCCAG CTCTTTGTAA TTGTGAAATC TGTACCCAAA CCTCTGGATT 60
AGAATCTCCA GTTGTCTACT GTAAATACTG GAATTACAGC AAAGGATATG GGGACTGGGC 120
TGCTTTTCTG TATTGTACAA GCACTATTCT AGATATTAAA GAAATTTAAC CGC 173

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

ATGGCGCCCT ACAGCCTACT GGTGACTCGG CTGCAGAAAG CTCTGGGTGT GCGGCAGTAC 60

CATGTGGCCT CAGTCCTGTG CCAACGGGCC AAGGTGGCGA TGAGCCANTT TGAGCCCAAC 120

GAGTACATCC ATTATGACCT GCTAGAGAAG AACATTAACA TTGTTCGCAA ACGACTGAAC 180

CGGCCGCTGA CCCTCTCGGA GAAGNTTGTG TATGGACACC TGGATGACCC CGCCAGCCAG 240

GAAATTGAGC GAGGCAAGTC GTACCTGCGG CTGCGGNCGG ACCGTGTGGC CATGCAGGAT 300

GCGACGGGCC AGATTGGCCA TGCTCCAGTT CATCAAG 337

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TGAAGCCCAA GNAGTNTGTG AAGACAGAGA ATGACCACAT CAACCTGAAG GTGGCCGGGC 60

AGGACGGGTC CGTGGTGCAG TTCAAGATCA AGAGGCACAC GCCGCTGAGC AAGCTGATGA 120

AGGCCTACTG AGAGAGGCAG GGCTTNTCAA KGAGGCAGAT CAGATTCAGK TTCGACGGGC 180

AGCCAATCAG TGAAACTGAC ACTCCAGCAC AG 212

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 265 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

CCTTTTAATA ATAATTCTGC TGTCTGCTGT GTACTAGAAC CCATGCCTAC TGCTTGGGGT 60

ATAATGTAGT AAATGTAGTA AAAACAATAT CCCCCGGGCG CGGTGGCTCA CGCCTGTAAT 120

TCCAGCACTT TGGGAGGCCA AGGAGGGCGG ATCACGAGGT CAGGAGATCG AGACCATCCT 180

GGCTAACACG GTGAAACCCC GTCTCTACTA AAAATA CAA AAATTAGCCA GGCGTGGTGA 240

TGGACGCCTG TAGTCCCAGC TACTC 265

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 333 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

AAAATGCTCA CAGTGGTCTT CTCTGGCCGG TGAGCCTACA GCTGATCTTG TCAGAGACAA 60

ACGTTAGTTT TACTGAGTCA CCCAGAGCCC TGTGCTGGTG CCTGAGGGTT TGTTCCATGG 120

GACAGTCTCC ACAATTCCTC TGGGGAAGGG CCACAAATCC CACAGTGTGT CCCAAGAGGG 180

CTGGAGTAGG CGGAGTCCCC AGCAGCTGTG GCATGACCAG CCATCTCTCT CAAAACAATT 240

GTTAACAAGC CTTCTGCAAG TTAAGGTTCC ACATGGTAGC CGTGGTACAG AGGCATTTCT 300

CTAGGGTGGG AGAGGCTTGT GCTCTACACC AGG 333

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CTCTGACTTT CCTGTGGNTT TAGAGCCAAG CTCAAGGTAG TAGGCCGTAG GGNCTTATTT 60

TATTTTCAAA CCCCCATCCT CAGAGCGCAG ATACATGCAG AGGCTTCTGC CAGGCTACCA 120

CGGGGCCTTA GTGGGAACAG GTTGAGACCA GCACTT 156

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 331 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CGAAAAGGGG NNNTATGGCC ATCTTTTATC AGAAAAAGTG ACAAAACGGG AATTTAAAAA 60

ATGAATTTTC NNTCTGACTT TATTTNNAAA TACACTTTCT TTTTTNNAAA ACCAATACAC 120

TTTCTTTGAG GATGACAGTA TTAGGAAATC CAATTNNACA AAAAATACTA CATCTAGTCT 180

GGGGTAGATA TATTTATTTT TGGTAACATA CATTAAGTGG CACTAATTAC ACAGTAACTA 240

TAAGGTAACT AACATGAAAC CACAGAACTG TAACTCTGCC ACAGCTGCAT GAACTTGGGC 300

TTTTCTGGTT GAGCCCATTT TCAAAAAACT G 331

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

AGCCACTGCG CCCCACCCCA TTTCGGTGTN ANCTCAGCTC ACTTCAACCT ACCCCTCCCA 60

AGTTCAAGTG ATTCTCCTAC CTCAGCCTCT TGAGTAGCTG GGATTACAGG GGTCTGCCAC 120

CACGCTGGGT GATTTTCCTA TTTTTAGTTG ACACTGCATT TCACCAGGTT GGCCAGGCTG 180

GTGTTGAACT CCTGACCTCA GCTGATCCAC CCGTCTCGGG GTCCCAAAGT GTTGGGATTA 240

CAGGTGTGAG CCACCACACC AGGCCCATAT TTTCTTTTAG ACATGCAGGC AATGTTGGTG 300

GGTTTGTCTG TTAAGA 316

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

GTTTTTCCTG CATCTATTGA GATAATCATG TGGTTTTTGT ATTTGGCTCT GTTTATATGC 60

TGGATTACAT TTATTGATTT GCGTATATTG AACCAGCCTT GCATCCCAGG GATGANGCCC 120

ACTNGATCAT GGTCGATAAG CTTTTTTGATG TGCTGCTGGA TTCGTTTTGC CAGTATTTTA 180

TTGAGGATTT TTGCATCAAT GTTCATCAAG GATATTGGNC TAAAAGTGTG CTGTATTCAG 240

GAAACCCATC TCACGTGCAG AGACACACAT AGGCTCAAAA TAAAGGGATG GAGGAAGATC 300

TACCAAGC 308

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

GGGGTTCCTC AATATGTAGG CGCCACTTTT TGTCGCTGTG CCGTCACCCTG GTCACCCCTG 60

TGTGCCCAAN ATCCCACTGT CTCTCTGGGT GTCCAAACTT CCTCTTCTTA GGAGGACACA 120

AGTCAGATTG GATTAGGGCC CACCCCAATG CCCTCATTTT AACTTAATTA CCTCCCTTTT 180

GTTTGGGCTT TTTAACTTAA TGACCTCTTT AAAGACCTTA TGTCCAACTA ACCTTTGATT 240

CTGAGGTATA CTCCACCTTA AGACTTTAAA ACACCAATTT GGACGGGATC TAATTCACTC 300

CATAACAATA ACAATAATGA CATCTTACAA CTTACTGCCA CCACCAAGCT TGCTG 355

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GGATGAGGTC GCCGGGATCG TGGCTGCACG CCACTGCAAG ACCAACATCG TCACAGCTTC 60

CGTGGACGCC ATTAATTTTC ATGACAAGAT CAGAAAAGGC TGCGTCATCA CCATCTCGGG 120

ACGCATGACC TTCACGAGCA ATAAGTCCAT GGAGATCGAG GTGTTGGTGG ACGCCGACCC 180

TGTTGTGGAC AGCTCTCAGA AGCGNTACCG GGCCGCCAGT GCCTTCTTCA CCTACGTGTC 240

GCTGAGCCAG GAAGGCAGGT CGCTGCCTGT GCCCCAGNTG GTGCCCGAGA CCGAGGACGA 300

GAAGAAGCGC TTTTAGGAAG GCAAAGGGCG GTACCTGCAG ATGAAGGCGA GGGAC 355

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

GTGTCTCTTT TAAAGAAAAC ATACTTTATT TTGGTCTAAA TTGTGAAAAT ACCCAAAACA 60

TTTGATAGAA ATTGAACTCT GTCAACAGTG TTATTTATAC TAAGATCAGG ACAGTTCCTT 120

GAGATCATAC TGTTTTATTA CTAAGTTTGG CCTTTGTTTT ACAAATGTAA TGTTCATATT 180

TATTTGAATT TTAAGATTGG TTAAATGTTA ATGAAAAGCA ATCCAATTGT TANTTTTTAG 240

TAGTGCCTTT TCTCTGTATG CCTTAATTTT ATT 273

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

ATTTTATTTC CTTACATCGA AGAAAATGTT AAAGAGTATC TGCAGACACA TTGGGAAGAA 60

GAGGAGTGCC AGCAGGATGT CAGTCTTTTG AGGAAACAGG CTGAAGAGGA CGCCCACCTG 120

GATGGGGCTG TTCCTATCCC TGCAGCATCT GGGAATGGAG TGGATGATCT GCAACAGATG 180

ATCCAGGCCG TGGTAGATAA TGTGTGCTGG CAGATGTCCC TGGNTCGAAA GACCACTGCA 240

CTCAAACAGC TGCAGGGCCA CATGTGGAGG GCGGCATTCA CAGCTGGGCG CATGAAAGCA 300

GAGTTCTTTG CAGATGTAGT TCCAGCAGTC AGGTAAGTGG AGAGAGGCCG GGATGAAGG 359

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TTTATNAAAG CAGTTAAACT TAGCATTAAA TAACACTCTT TAAATGGTAC ACCTATGAAG 60

CAAGAGTTAA ATATAAACCC AGTCTAATCC TGTACACTTG TGATTAATTG TGACAATCTT 120

AAGTTGCTCA CTTCTTTCCC ATTTACCAAT TCAGAGAAAG CCCGTTTCCT GTTTTCTCCT 180

CACCACTTTG CCTTGGCATC ACACCAACCC TGCCTCGGGC TTCAGCTGCA GATCCTCCCC 240

AGCCCCTCCT CCCAGCTGGG CTGACTCCAG TCCCAGCCCC AGTCTCCACC AACTGAGCAG 300

CGTACGCAGG GTTGTGTCTG GCTTCCAGCA TCTACCAACC CTTCAGAGCA ACTTCCAACA 360

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

TCTCAGAGGG GCTCTGGGGG TGATTCAAGG GGGACTTCTA GCTTCTCTCT GGAACCCTTT 60

GTCCAGAGCA AAGCCAGGTT TCCAAGGTCC CCACGGCAAG GCTGTTGGGT GCTGGCAGCA 120

AGAGGTACAC AGCAGTTCTC CCAGCTCACA GCAGTGACCT CAGATCTCCA GCAGCAAGGG 180

CCGCACTCTC GTGCCCACAA GGGCCTTGCA GAAATNCTCC GGTCCCTGGG NCTCCCCCGG 240

GAGGAGGGGC GGGGCTCCTG CCTGCAGTGA GGCCACAGCA CTAAGCGGCT TCAGTCACAT 300

GCTTTTCAGG TGAATCACTC CAAATTCAGT GAGGAGGCCC ACGACAAGGA AGTTCAGGTA 360

CAAG 364

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 488 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

TTTTTTTTTT TATATTTTAA ATGGAATTTA TTCTATCAAC TGCCTGAGAG GACACAATGG 60

GGGAGGGGCT TCGGACCACA GCAGGAGCCC CGACTGCCCA CCTGAGGGCA GGGAGAGCCT 120

GACCCCATTG GCCCAGGCCC TGGCTCTGTA ACCATTAACC TCTTCCCCCA ACTAACACCA 180

ATGAAAACAC CATTCCACGT GACTGGGCTG TGTGTTTGCC TCTGTGACAT GGGGACCCCT 240

GACCCTAGGG GTCTCGCCTG AGCCAGACCT GAGGGACCCA CCCGCGTAGG ATGGAGGAAG 300

GTTTAGGCCT CCCTTTTGCC AGCCAACGCC GGGGGGTGGG GCAGACCCTG GGAGTGGGCC 360

TTACAGACCA GCCACAGGTA TTTCTTAGGC AATTTGACAC ATTTTATTAC AAAACCAGTC 420

TACATTCATT CCTAAAAGGG TCATTTTCAG TAAAA 455

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CTGATCTGAC AGGAGGTGTA GGTCAGGCAG TAATGGAAGT SATGGGGAAC AGCTGTAAAT 60

ACAGATAAAG CTTTACTCAC TCGCCCACCC ACTGCTCATC TCCTGCTGTA CTGCCCAGTT 120

CCTAACAGAC AGCAGACAGC TACTGGTCTG TSGCCCAAGG GTTGGGGACC CCTGACATAG 180

ACTAAACAAT TCACAATGTT TATATTAAAC AACTTATTCC AAGTTTCCAT TTTAGACTCT 240

GGAACATCTG ACATGGTGAA TCCACAGGTA GTAAATSGGA AGGGAGATAA CAGACAACTT 300

GACGGCCGTG GAAGACGCAC TGGGCGGGCA CTGGTGACGG GTCTCGGGAC AGACTTCACA 360

TCTCCAGACT GGCACAGTGG GCTCACACCT GCCTCCCA 398

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

ATCAGTGTCA GTGTCTAACA GAAGGGTCTG TTAAGGATGC TTCTGATTTA ACCAAAAGAT 60

TAAGCTTCAG AAACAATCTA ACATACTCAA AGGAGCACCA AATTATCAAC CGGCTACAAG 120

GATGCAAAGG ACCTAAACAA CAGATGTCAA AGGGCTTGTA AAAACTGGAG CCAGCAACCA 180

TTCCACTTGA AGGAATCCAT CTCAGGGAAA TGCTGGAATC CACACACAAA AGCAGGTGTG 240

CAAATAATCA CTGCAGCACG CCTTCTAATA GTGAACAACA GAGGCAATCC AAATATCCTT 300

CAACAGGGAA CTGAGTAAAT ACCAACTATG GGCATATCCA CATAAGGCTC TCTGCAGTCA 360

TTAAAAAAGGA TTGCACTTAC ATGCATGTCT GCCATGGAGG TCTTTCAGGC CAATGGTTGC 420

ACTCGGAAGG GCAACCACCA ATTA 444

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TGGGGCCCCA ACGGAGACCT GGGGATGCCG GTGGAGGCGG GAGCGGAAGG CGAGGAGGAC 60

GGCTTCGGGG AAGCAGAATA CGCTGCCATC AACTCCATGC TGGACCAGAT CAACTCCTGT 120

CTGGACCACC TGGAGGAGAA GAATGACCAC CTCCACGGNCC GCCTCCAGGA GCTGCTGGAG 180

TCCAACCGGC AGACACGCCT GGAGTTCCAG CAGCAGCTCG GGGAGGCCCC CAGTGATGCC 240

AGCCCCTAGG CTCCAAGAGC CCCCAACCGG GACCCAACCC TGCCTCCCTG GGGCTAAGCT 300

CTGGCCTGGG GCACTCACCC CCTGGCTTAG ACAACTTCTC AAGGGCTTGG CCTTCAGGGG 360

ACCCTTGTGG GTCTTGCCTT GCTGGGGCCA CCTTTTCTTG CTTGGGGCTT CCCCTTTGGC 420

CTACCTTGGG GCCAAGCCCC TACCAACTTT GGATTGCCTT CTTGGGGGCC AA 472

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GTNGGGGCCA TGTGGCTGAT TTCCATCACC TTCCTTCCAT TNGCTACGGC GACATGGTGC 60

CCCACGACCTA CTGCGGGAAG GGTGTGTGCC TNCTCACTGG CATCATGAGA GCTGGCTTTA 120

CCGCCCTCGT GGTGGCTGTG GTRGCTCRCA AGCTGGAGCT CACCAAGGCT GAGAAGCACG 180

TGCACAACTT CATCATTGAC AGTCAGCTCA CCAAGCGGGT AAAAAACGAG GCTGCTAACG 240

TTCTCAGGGA GACGTTGGCT CATCTACAAA CATACCAGAG CTGGTGAAAG AAG 293

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

TTTGAAAATT TAGAGGATAT TTATTTCTCA GGAAGGTGCA CAACAGCTGG CAGGCACTGC 60

TTTCCCTGCT CTAGGGGATT CCTCTCTCCT TTTCCAAGAA ATCCCCTCTC TTCTTAGAAG 120

TGCCCATGGG AGGCTGGGAT GTGAAAAGAA ACCATACACA ACACTCCAGA GCCTTAAAAA 180

AATAAAGCAA CAACCTCCTC CACACGAATA CACTTACAAA ATAAATAGAC GGATAAAAGA 240

GAGGCCACGT GCCTCCCATC CCGGCTGTAG GGCTGCTTGG GGATAGTGGG GCTGGGTGGC 300

TCGGTCCCAC TTCTCCCAGC CAGGATGATC CAAAGGCTAA ATGGGATGGA AGGGCCCTGG 360

CTTTCAGAGA GAGGGTGGGG CAGGCCTCTC CTGGTACTCA GCAGGGAGGA CACTGGGGCA 420

CGGGTAGGGG TCCAAGGGCC ACTTAATA 448

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 551 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

GAGACGGAGG CTCGCTCTGT CCCCCAGGCT GGAGTGCAGT GGCGAGATCT CAGCTCACTG 60

CAAGCTCCGC CTCCCGGGTT CACGCCATTC TCCTGCCTCA GCCTCCCGAG TAGCTGGGAG 120

CCAGCGCGCC CAGCCTAAAA AACTTTTCAA GTCAATATTA CTACGATTTA ACATTAGAGT 180

GTGGACATGT GATTTAATCG CTATAGCTAA AATACGTCAA ATATACGTTG TCATGTGCTT 240

GAACATGATG CTAACCCTGA CAGGATGAAG GAAAGTAATA TTCTTTCAGT GTAGTTCAGG 300

AGAGCATTTG TTTTCTTTTC TACCAATTAA CCCATCATTG CTTTTAAACA ACCATCTGAA 360

GGAGCAGAGA GGCAGGGTAG AAGACAGAAG GGGGTCTATG TGGGTACTAA AGATGTTTCT 420

GTTTTGTAAT ATTGTGTGTG TGTGGGTTTA TGGTTTGCTT AAGGGATCAA AACCTGGAAA 480

AAATGGGATT CCAGGAATGG CTCTGTTATT TTTGCTGGGT TCCAGCTTGT AATGCCTACT 540

GCCTTGGTTC A 551

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

CCCCACCCCA AAATCAAAAC TGAAGGTAGT GTCAGTGTAT ATATGGNGTC CCTTGTGCTG 60

AAAGTCAAAG CAGCTTCATT TTGGGGCCTC AAGAGCTCCA GCTCTGGGCT CTTCACCTCT 120

AAGCCCATGG GCAGTGCCCG CCCAGTGGTG TGTATAGATC GGAGGCTGAG GGCCTCACCC 180

TTAGCTGAGC TGTCGCGTGC TGGGGAGCCT GTGCAGGAGG GTACAAGTAG GAAAGTGCCA 240

TCTGCATGGG AAGAAAAATG CAGCGTCCTT GGTAGTGCGG ATGGGGTCCA GGAGACCCAG 300

GGAGCTTGCC CAGAGGGACC TGAGTGGCAT TCCTGTAGGA AAGCAGCCCA GATCTTGGGG 360

CCGTAACGGA TGTTCTGGAA GTTTTGACTT TGAACCACCA GGTCCCATTG TTAACAAGCT 420

TCTTGA 426

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

TTTTTGGGTT AAAAAGGCCC AAAACTTTAT TTAGTTTTCA GGGAAATATA AGATGCATGT 60

AAACATAAAA TACAAAACAA AACCCAAATC TTACAGTCTA GAAGCATGCC AAGACAGAGC 120

ATTTTCTGCA GACCAAAGAG TCCCGTCAAA GTGATAAAGG ACACCTGGAA AGTGGCAGGC 180

CAAGGGGCTG GTCCCTTCCC CAAGGGCACT GCATTTTTGT GATGAGATTA AAAACAAACC 240

AACTCCACTA TTAAAAATGC TAGAAACATG GGATAGTTTA GCACCACCAT TGATTCTGGC 300

AAATATTTCA GCACTCACAT CGACTGCACT GAGTTTAATG TCCTTTCTCC AGTTTCTCTG 360

CTGAGGAGGG AAGGAGGGAA ACCTGGGCGG AAGGGGCTCC TCCTGACCCC ACAGGGCCAC 420

TAGGAGCTTG GAGG 434

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 276 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

AGGAAGTGTT AGCAAATGCT ACCATGTGGA ACACTCAACT TTATTTGCTT TATTTATATA 60

TTTAACAATT CTAAAGTATT TACTTCTTGC TTTGACAAAA AATGAAAAAT ATAGGGGCAC 120

TGACTGACTC CTCTTTAGGA GAAAAGGGTT ATATGTACAG GTATGGAGAG TTAGGGTTCC 180

CGCTTTAACA AAGGCAAATA TTAATAAAAA AGGGCTTCAT GGGTCAAAAA AGGGCTAAGA 240

GGTGCAAGCA TTTATTCACA CTGTACATCC GGCCCC 276

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 554 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

ATTTCTTTCC TTAATCATAT CTGATGCTGG GATGTGGGTA ACCCCAAACT GAAGGCAGCT 60

GCTAAATCTC AAATGCTAAA AAAATACTGC AATTTTGACA TCAGTGAGTC AGATCAATAC 120

ATCCTCTGGG GCTGATTTTG CTTCACAGTT AGGATGAGCC ATCTCTTAAG CTGCAGGCTC 180

AAATGGGATT AACTGAACTC TATACCTGGG ATGGGCCATG GACTGAGCTG TCCATGCAGA 240

AGGACCAGGC TGTCCATGCC TTCCCTGCCC TTTTACTCAC CACTGCACAG CAGCCCCAGT 300

GGGCCTACTG CACATGTCTA GGAGAAATCA CTCTAAGAAA ACCAACAGGA ACAGGCTTTA 360

GGCAACAAGA GACGTCTCAC TGCATCTCCT CCCACGTCAG AACTTGAGTA CTGGGTCTTT 420

GCAGCTCAGA GCATTCCTCC CTTCCCTTTC CTGCCCGAAA GGCCTGCCTT TTCCTGAGAC 480

ATATGGCACT CCATGCTGCA AGTTTCAAGC AGATGCAGGT TCTTATGGGG CTTTTTGCTC 540

AAAGAGCTTT GGTT 554

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

CACCTAAGCA GGTAGACATC CGCAAAGTCA GATGCTTTCC AACATGACAC CTGAACATCT 60

TCCTTTATGC AACACCCAAA CATCTTGGCA TCCCCACCCC AGGAAGTGCG GGGAGGAGGT 120

TATGATCCCT GGGCGCTTCG GCAGAATGGA GAGCTGAGGT GTCCCTCCCC TGCTAGTCAC 180

CTACCAGGTG TCTGAGCAGC TGCATGCTCC CTGGCTCAAG TGGGCACTGT ACCTTTTG 238

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 244 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

ATCCAGGCTT TCATTTCTAG CCAACCCTCA AACACCACCA ACTACAAAGA AAATTTAAAA 60

GTCTAATTTG TAACCTTCAG ATAAGTATAA ATTAGTTTTT TCTAGGCTTT CATTATTTGG 120

CTTCTTATAC AATCTATCTT GTAAAGTACA TTCCTCTAAA TTTACATTAT CTAAAATTAA 180

GGCTAAGCAT TATTTAAATC ANTTAATCAT ACAATATTTT ATGGCAATAT GCACATATTT 240

ATAA 344

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

CTCAGCGTAT CATAGGCGTG CTCACCCTCC TCCCCACGCT CCCGCCCCGC AGGCAGGTGG 60

TGTAGGATAG AGTGGTGCAT GAAAGGGGGG AAGCCCGAGG GGCCCGCTGG GAAGGGTGCT 120

GCCCCGTAAA GGGCATCCCA CTGGCACTGT GCCTCANCTG CCGCTTTCTG CTTCAGCTCA 180

GCCAGTCGCC GCCGCTGCTC TTCAATCACT TGTTGTCCCT TCTGCTGCAG AGCTAGTTGG 240

CGCTTTGGTC TCGATGTCCT GCAGTGTGGC TGCCAGGTTG CAAGGAAGGC TGCCCGGTGC 300

CATTCTGGGG GTGAGTAGGA GCGCTCTTTT 330

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 281 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CCTCTCTCCC TTCGGTTCTC CATTTACCGA GCCACAGTAT TTCTTAAAGC TCGTTGGCAG 60

CCTGCACCCT GCTTATTCTT GGGAGACACG AGTTTGCATC CTATTACAAC CCATAGTTTT 120

TGCATAACCA TGGTGAGAGG AACCATCCTT CCCAATCCCA ACCTCAACCA AAGCTTAGAA 180

AAAGTGCCAT CNTTAACCTT TCAGAATCAC TCATAAGTAA ATCCTATAGC AGTCTCTGCT 240

AATGCAAATT TCAATGTGTG CCCGCTTATT AGGTGACTTT T 281

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 266 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

CTTTTAATGA TGTGGTTCTG GTGGGATTTA TAAATCGAGA TGAACCCCTG GNAAGATGCT 60

TTCCTNAAGC ACAACCCACA CATTGGGTGA CCATTTTCCT TTCTTCCTCC TTCTCTGGGT 120

GGCCCGAGAC CTGTAGGACC TTCCCTCCCT TTAATTTTCT TTAACCCCCC TTNTGAGTCC 180

TGAGACTCCA TTCTTCCCTT GTCCTGACCC CCT A T ATATA TTTCTCCTCC 240

TCTTAGTTTG CTGTCGCGTC TGTTTT 266

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

GTCTGGTTCT ATTCATTTTG TAGTTGCGAG AAAAGGAATG AACCGTGACT ATGGCAATTC 60

ACCGTGACGT GTGATAATTT AGTTTGCTAT GAGTTTTCAC TCTTAGGTAA AACCTAGTTA 120

TCCTAATTAA TAATTAGTTA TGGATGATAT AGTAATTTTT TTTTTTTTTG ACTGCGTCTC 180

ACTGTCATTC GGGCTGGAGT ACAGTGGCTG ATCACAGTTC GGTGCAGCCT CGACCTCCCT 240

GGGCTCAGTG ATTCTCCTGC CTCAGCTTCC CAAGTGGCTG GGGATTATGG GCATGCACCA 300

TCAATGTCTG GCTAATGTTT GGTGTGTTTT TTTATAAAGC CAAGGGTTTT GCCCATGNTT 360

CAAGACCCCG GGGCTGGTCC TTGAACCTCT TTGGGGCTTC AGGCAAGTCC TCCCACCTTC 420

GGGCCTTCCC AAAGT 435

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

TTCCCTTCCC AAGGACTCGA CCTGAGAACC GCCATGTACT CGGAGATCCA GAGGGAGCGG 60

GCAGACATTG GGGGCCTGAT GGCCCGGCCA GAATACAGAG AGTGGAATCC GGAGCTCATC 120

AAGCCCAAGA AGCTGCTGAA CCCCGTGAAG GCCTCTCGGA GTCACCAGGA GCTCCACCGG 180

GAGCTGCTCA TGAACCACAG AAGGGGCCTT GGTGTGGACA GCAAGCCAGA GCTGCAGCGT 240

GTCCTAGAGC ACCGCCGGCG GAACCAGCTC ATCAAGAAGA AGAAGGAGGA GCTGGAAGCC 300

AAAGCGGCTG CAGTGCCCCT TTGAGCAGGA GCTGCTGAGA CGGCAGCAGA GGCTGAACCA 360

GCTGGAAAAA CCACCAGAGA AGGAAGAGGT TCACGCCCCC GAGTTTATTA AGTCAAGGGA 420

AACCTTCGGA GATTTCCACA CTGACCAGCG AGAGAGAGAG CTTTAGGGCC A 471

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GCCTTTAACA TCCTCTGCCA ATRACTGGCC TCAAATCACC AGTGGAACCT TTTCAAAAAA 60

TACACCATTG GCTCTATGTA GTTCTACTGA TCTRAAATAT CCACGTGTGG GCCAGGAGCA 120

CTGGCTCATG CCTGTAATCC CAGCATCTTG GGAGAGCGAG GAAGGAGGAT CATTTRAGCC 180

CAGGAG 186

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

ATAAAATACT TAGGAATATA CCTAACCAAG AAGGTGAAAA ACCTCTCCAA GGAAAACTAT 60

GAAACACTGC TGAAAGAAAT CATAGACTAC ACAAATACAT TTCATGCTCA AGGATGGGTA 120

GAATCAATAT TGTGAAAATG GCCATACTGC CAAAAGGGAT CTWCAAATTC AACGGTATCC 180

CCATYAAATA CCACCATCMT TCTTTACAGG NTTCGGAAAA GGAATTCTAA AATTCATATG 240

GGACCCAAGA CGGGGCCCGC ATAGCCCATG GCCGGCTTAG SAAWAAGGGA CAAATCTGGG 300

AGGCCTT 307

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 184 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

CCAGGTTGGA TGGAGTGCAA TGGCACGATC TCGGCTCACT GCAACCTCCC AGGTTCAAGC 60

AATTATCCTG TCTCAGCCTC CTGAGTAGCC GGGATTACAG GCACGTGCCA CCACACCCAG 120

CCAATTTTTG TATTTTTAGT AGAGACGGGG TTTCACCGTG TTAGCCAGGA TGGTCTCAAT 180

CTCC 184

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

GCNGGAGGCC CTCAAGGGCC AGGAGCTATT CTACACCCCC GAAATCGCCG ATTTCAACTC 60

AGAACTMTTC GMGNAGACAG CCAGGAGCAT TGAGAGCACC CTGGACGACC TCTTCCGGAA 120

TTCAGACGTC AAGAAGGATT TCCGGAGTGT CCGCTTGCGG GACCTGGGGC CCGGCAAATC 180

CTTCCGNNNC ATTGTGGATG TCCACTTTAA CCCCACCACA GCCTTCAGGG CACCCGACGT 240

GGCCCGGGCC CTGCTCCGGT AGATCCAGGT 270

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 529 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

CTTGCAGTAC ATAGCATTGT TATTACTGAT AGCTTTATAA ATCTGCCAAA TAACATAGAA 60

TGTAGCCTCA AAAGGATGGT CGAGGGTTCG CAATCTTTCT TTCTCCACCC AGTGGTGTGG 120

AGCAACTCTG TGCCTTAAAG AGGGCACCAT GGAAAGAAAC AAAAAGGAAT CTCTTTCAAA 180

ATGCTGGAAA TTAGGCTTAG CTCACTACTT TCAGGATAAA GACAACTGCA TCTAATTAAG 240

TCCACTCCAC ATTTCTTTGG ACTCTAAGTA TTCTGCACCT GAAGGCTAAA TTGAACTGGC 300

TCAGCCCTAT CTTTTTTGCC ACATCTTTAA TTACAAATCT ATTTCTTCTT CCTTTCATTT 360

ACTTCTCTTC TCTTAAGTAA GAAATGTGGG AAATGAGACT GGCAGTTTGG TTTGTTTGCA 420

TGTGGGTGTC CATTAGGCGT CTCATCCTAT GGCCCTTTTT GGAAATGTTG CCTTCCTACT 480

ACACACCTGG GAGGTTTCCC CAAGGCTCAA CCTTTTTGCT TCAGGTAAA 529

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

GACGGTGGCG ACGCGTGCAC CGGGGATGTG TCCTGCCACC AGAGGAGGTG TGCGTGGCGG 60

GGAGCAGAGG GGCTTTGTTT CCCAGGTGAA GGTGCGGCTT CTTCACTCTT AGAGGTGCGT 120

GTGTGGGTGG GGGTGCTTGC TGTTGAGGTT TATGCCTGTA ACTGACAGCT GTCCCCCAAG 180

CCATGCTGGC AGTGTGTAGG TGTCGTGCCG GCCACCGCAG AGGAATCCTC TGGGCTTCTG 240

TGGTTCAAGT GGGGCCCAGC GCAGAGCTCC ATGAGTTGCT GAGCAGCCAG CCCTTCAGCA 300

TCTCCTGGGT TTTGGCAGCA GGAGGCGTCC CCTTGTGCAA TTCAGGGGGC CGTGGGGGCT 360

GGGGGCACTC GTAGCAAGGT AAAGGAGCCC CTGCTCAGGC CCTTGTTTGC TCCCCTTTCT 420

TGCAAGAGGG GTAGACG 437

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

GGCATTGTTC TGGTGGGTGT GTCACGCTCC CAGAAGACTG AATTTATGGT AGGATCACTC 60

GCAAGGCCTT GTGAAGGAGT CTTACCTAAA ACAAAAGAAA TATCAGGGAC TTTTGTTGAC 120

TATTTACAAC TCAGTTTTAC ATTTAAATTC AGGCAGTGTT AATATGCCAA GGTAGGGAAT 180

GTGCCTTTTT CAGAGTTGGC CAGGAGCTCC TGGCTGGGAC ACGGAGAGGC AGGTGTGGCG 240

TAAGGCCTCA CTCCCGGCTG TGAAGGTCTC TGATCACACA GAAGCAGCCC TGCCCAGCCT 300

GGGTCATTTG CTGTCCGCTT TTCTCTGTGA CCACAAGCAG CCCTGAACAA CCAGTATGTG 360

TCTTCTTTCT CCAGATAGTG AAAAAGGGTG TCCAGATAAA CCCACCTAAG TGAAATGGGC 420

CATCCTCTAA ACTGGGGTAC CTCACTGCAC AGGTTCTAGG TAGGCTTTCC ACTTAATCTA 480

ACTTGAGGCC TACAGGTACC CTGTAAAGTT AGTGGGGCTT GTCCTTGATT GTGG 534

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 279 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CAGTTTGGAC AAAGTAGCAT AGTGAGTTTN TTCCTACANT GACTTTCGGA GAAGTTNGCA 60

GTTTCTGGCA AAGTGACGCT GGGCTGTTTG AAAAAGGCAA GCTTAGCCTA GGCTGCCATC 120

TTAAAACATT TGGAGGCTGT AGCTTCCTCA GGATCCTTTG CCTGTGGTCT GGTGGCCGGC 180

AGTGCCCCGT CTAACAGCTT TTAACTCTGC ACTTAGTGCC TGAGCACCTA TGGCTGTGAG 240

AGATGCTAGA TACAGAACCC TGTCCTGTAC CACGTGGGG 279

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 518 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CAAATATTTA ATGCACATCT TCCTTCTTGG TCTCTTATAT CTCTATCCCT TTTTGGTCC 60

TTTAGGAGAA TCTGTACTAT TTCAGCATGT CCTCCTCCAG CAGCAAAATG AAGAGGAGAA 120

CTAAGTTGTC CATTTAAAAG GTTTGGATTG CACTTTCCTT TCTCTAACAA TATGCGAGTG 180

GCCTCAACTT TTCCATACCA GCATGCATAA TGAATGGGTG CCCAGTGGTC ACTATCTAAC 240

TGGTTGACTG AAAATCTTTC ACTGAGAAGA CGGCTTAGTA ATTCTGAATC TCCTTCACAG 300

GCGCTTCGGT GGAGAGGAAA ATCATCTACC CACTGTCGTT CCTTGTCTTC TGTGACACTG 360

CTCATGCTTC TCTGCCAGTT TTTCCTGTTT AGGGTATTTG GATTTTTGAG TAGTCTGGAG 420

CTCCTAGACC CAAGTATGGA TTTATTACCC ACTTATCTAC CCGATTTGTA TACTGAGGAT 480

CCTATCCAAC AAAGGGTGTA AATCCAGGAT CCGCCTTC 518

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 266 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

GATTGCAGGC ATGANCCACT GCGCCCAGTC GAGTGGTAAT ATGTTMAAAG GAAACCTTTT 60

TCTGAGCAGG TCTCAAAAGA GAGGTTAAAA TACTGAGTAG ACCATMCTGT AAACAGATGT 120

MCTGTTATYC GGGCTTTCAT ATTCCATTTA TAAAGCACAG GCAGAGCTCA GAGTAGATTT 180

AAYGTAACTC TGAAGGGCAC TAGGATTTTC AGAATGGTAA ATAAGCATTG GCTTCACCTT 240

AAATYCAAAT CTGCATTGGG CTTGTA 266

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

ACCTCGCTCA CCGCTCTGAC CACCGACAGG CAGAGCAAAG GATGCGGGAG TTGCCTCTGC 60

TGCCCATCTA AGGGGACGTA GGCAGAGAAG CAAAGGCCTC TGCTCTCCCT CCATCCATCC 120

CGGTGTGCTG GCCCCAACGG AACAGGAGTC CTTCAACTAT TGCCTGCCAG AGACCCAATT 180

TTAGGGACTG TAGTCTGCAT CTGGATGAGC TGGGCTGTAG ATTGAAGTCT CAGAAGCAGG 240

GAAGGTTGGA AGGGGTAGGG TCCCAGAGCC CATGGAGTTA TTGCTGAGAA GATATGCAGG 300

GGACACATTT CCCAGGGGCA GAGTAGAAGC CCTGGGCCTT G 341

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 234 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

GTGAAGGGAG TTGCAGAATC AAATTGCTAC ATAGGCCAAA CAAAAAAGAA GGCTTTTTCA 60

AAAAACATTA AATTCACATG CAGTCTCAGA GACTATTTAG GCAAAGTTCA AGTTAGGAGC 120

TTTTAGGATG TGGGANTAAA ACTTTAATKG GAGGGGAGGG CTTGCTTCTG GAGAAGGAAG 180

AAGCCAGACT TGTTAGACAG TACTCTTAAC TCCTAGCCCA GCCTAGCGTG CCCT 234

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

CAACTCAGGT TAGCAACTGC AGGAAAAACTT TCTTCATTTT CACTGAATTT TAAAGAGAGA 60

ATGCTGTCTC TATTTCTCAG AGAAACTTAG GTGAAAAGTA AAAGAGAGGC AAAATCTCTT 120

TCCTTCATGA GATACTTTTA TTTTTATCTC TTTCTCTACT CATGTGCTTA ACTGGTGAAA 180

TGATTCTGTA GAAATAGATC CTTCTGATTC TGCATCTCAT TTCCTTATGG CAACTACAAC 240

AGGAGGAATC CAGCTGGAAA TGCCACTAAC CCCACATCCA GCACCTGAGA GAGGAAGCCA 300

GTCGGAGCGC CGTGCTGGGC TCACTCACTC TGGGCCTGCG CACTGGGGTT GTGG 354

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 373 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

GTTTTTGCAA CACTTTTTTT TTAAGTTATT GGGTGCAAAA TCCCAAACCA GGATATGTGT 60

ATGTCTGTGT GTTTATGTTT TTNATTTGAC CCTCCCCTCT TTCAACCTAC CCCCTTTTAT 120

ATCTAATGTA GAAAAACCGA AATTGAATCT GGAAAGCAAA GTGTTGTATA TACTTCCGGT 180

AACAATCATC AACACAGAGC CGGGCTGTCC AGTTGTTTTT GAGACAGACT CTCACTCTGT 240

TGCCCAGGCT GGAGTGCAGT AGCATGATCT TGGCTCACTG CAACCTCCGC CTCCCTGGGT 300

TTAAGCGATT CTCCTGCCTC AGCCTCCCGA AAGTAGCTGG GATTACAGAC GCCTACCACC 360

ACAATTCCCC TAA 373

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 262 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

CCGCACCTCG GCCAGAGGCG GCTGCAGCAG CTGCTMCCTT TTCCCTGCCG CCGCCTCTCC 60

AGTCCCTTTT TTAATTACCA CTCCAMCTGC TGGGAACGGG CGAGAAAGAG GAGGAGGCGA 120

GAAACTCCCA CCGACCCACA GAGGGAGCAT GATTTCGGCA ACTTCACCTA TCATTCTGAA 180

ATGGGACCCC AAAATTTTGG AAATCCGGAC GCTAACAGTG GAAAGGCTGT TGGAGCCACT 240

TGTTACACAG GTGACTACAC TT 262

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

GGAAAAGCGG GACCCAAACA GTGGTGCTGG GGAAATTTTT CCCTGTCCCC TTTGGAAGGC 60

TGAGTGGGTG ATGCAGCACA GGAACAAGGC TTGGACGTCA GAGGTCTCAT CTTCACTGTN 120

ACAAAGCATA AAGGACTTGG GGTTGAGCGT GTGTNTGGGC TCAAGTGACC ATGCAAGTCC 180

TGTCACCTCC TTCCTAAGAC CCCATCCTTC TCCCAAGTCC TCCACAAGAG CTACCTTCTT 240

CAAAACAATA ACAGAAACAC ATCAAGCTTG GGCGTCACTG AATTCAAGTT CTGATTTCTC 300

CCGTCACCCC AGCAACAGTG CCCAGTTTGA TTGTGACACT TTGACCCAGC ACTTGGTTTT 360

GAATGTTCTT TTCGGCTTGT ACCG 384

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 324 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

CTACATGGAA TCATAAGTKT TCCTAAAAAA GGAAGACAGA TTTGAAGACA GAGGAGGAAG 60

GTGATGTGAT GATGGAAACA AGGGGAGAAA ACGCAATGTG ATGTGGCCAC GAACCAAGTA 120

ATGAGGACAG CCTACAGAAG CTGGTCAAGG CAAGGAAACA GATTCTCCTC TAAAGTCCCT 180

GGAGAGGGCC TGGCCATGCT GACACCTTGA TTTTKTCCCA GCAGAAACTC ATTTTGGATT 240

TCTGGGCTCC CAGAAAAGTA AGGGGGTAAT GTGCTGTTTT ATGTCAGGTT TKGGGTAATT 300

TGTTTATTGC AGCCATCGGG AAGG 324

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

TTTGCCTCCT TGCTTGCTTA TCCAAGCAAG GGTGTGGTGA CAATGACCTG ATCGGGGTTT 60

AACGCCGGCT CTGTCTGCTC ACCAGACCTG GGGTGCTGAG CTCTGACCAG CCTGGGCAGC 120

CCAACCCACA GGAACTGCGG TTTCATAGGT GGGTCTTCAG GAAGGGGTGG AGGCTTTGGG 180

AGTGGCAGCT CCCCGCCTCC CACCACCCCA AGCCAGAGAA TGGGGCAAAC TTGTATGCAT 240

GGCTTATCTC TAAATTACTA ATCTGCTTCG GACCAGACTC ATCTCTACAG TATAGAGTTA 300

GAGTTATTGC TTCTATGACA GGTGTTCCAG AAGCCCTGGG TGGCTTTAAA GTCTG 355

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

CAGTTTTCTG AGTTCCCGTG TGCTAGACTG GCCAGAAGAG AGGGTCTGGG GCCTGGTCAC 60

TGGGCCACTC TCTCCTGTTT CTGGCCTCTT CTCCCTTCAC TCCCGTCCAG TCTGGTTTTG 120

AGAGCAGGGG CTGTTCTACA GCACCTCAGG GAAGGGAGGA GAGATACCTG CTGCTTCCAT 180

TGCTTTTCCC TTCCTGGAGT CGATGCCTTT CTAAGGGTTG GAGCTGCTCC TTCCAGGGGC 240

GGGTCAGTTT CCCAGGCCAT GCCGGGGGTG GCCATCTATG CTAGGGCTGG AAGCTGAGGC 300

TGGCCGCCAA CTGTGGGGCT GGGGTGGGGG TGGGTGG 337

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

GGAATTAGAT GCTCAGGTGT CCAAGCAGGG ATAAGCACAG CCAAAATAAA TAACCCGCCC 60

AACCCCCATT GTCACTCTGC TGCAACACGA CACAAAGGTT TAAACATCTG GGCCCALAGA 120

CTCTGGGACC CTTCAAGCAA GTCAGGTGGA AGAAGGTTTC CCCACCCCCC ACCAGGCCTG 180

TTTGTCCCAG GTTGCCCTAG GATGGAGGCA GTTCAGACCC TGGGTCACTG ATGCTTGATA 240

GGAAGATCTT TGATATCAAT GGCCTAAGCT CTGCTCAT 278

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 368 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TTTTTTTTTT GTTTTCAACA AACTTTACTA AATAACCCTG GAAAGGCAAT GAACGATCTG 60

ACAATTTAAG CTCTAATGAT TTAAAGCTCA GCTAGAAGAA AGTGAGGCAT GACATATACT 120

GTCAACGGAG GGTGAAGGAG GCAGATTTCT GGAAATGCAA TGATCCCACA CATTTGCTTC 180

AAGGAGAAAC CTGCAGACAT ATTTTCAGGT CTTGCTAAGT AACAACTGTT TATTTGTAAT 240

CAATACATTT GGGGAAAGTC TGCTATGTAG CTAAGGTCAC TGTGACCACA GACCAACAGA 300

TGGAAAGGAA AAAGGCACTG GACCAGCAAG GAAAAATACA TCCCCATCCT CAAAAGAATT 360

TTAAGGTG 368

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 367 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

TTGTGAAATG GGCCTGGGTA GATAAGGAAA AGAACCTCCA AGAGGTTAAG TGATTTGCGG 60

ATTTGCCTAA ATTATACAGA AGAGTCAGCA CCAGTGCCCA GGCCTTCTGA TTCTTAGTGC 120

AGTAAACACT AAGCACCATC ATTCCATTTC ACCACACTCC TGTCTTGCTG TTGTCCTCAG 180

CTAAGAAAGC CTACCCCTGA GTTACCCTCT TCCATCTTAG AGCCTTCCTG CTCGCTGTCT 240

GCCCCCCTGC GATGGGGACT TCTTTGGCCC TTCTCACCCA GCCCAGCCTC TGCCCGTTTT 300

CCTTCTCCTT TCCACTGCGG CTGAGCTCTT TTCTCCTTCC GAGAAGCCTT TCCTTCATCT 360

TTCCTGG 367

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 366 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

CCCAGCGGGC CGCCTCCCTC CTCTCTCCTC CATAGGTGGG GGTTGTGGGC CTTCTTTTTT 60

TTTTTGTCTT GGAGGGCAGT TAAACTTCTC CATTTGCCTC TCTCTTCACA CCCAAATGCC 120

AAAGGACACT TTTCCTTTCT TTTGTGGGTA GTTGCAAAAA AAAAAAATTC CTATGGGTTA 180

CTGCCACTTT TAAATACTTT GTAACTTAAA GGCAAAGTAG TATGTCACTG TTTCTTTTCC 240

CTGTAGTTTA CTTTTGAGGT TAAACATCTT TCCATGTCTT TATTGGTCAA ATACAGTTCC 300

TNCTTTTGTA CAATGTTAAT CCTAATATGG ACCATTTTTC CTAATGGGAT TACCGATTTT 360

TTTAAA 366

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GTTATTCTGG CAAGTGCTTT CAGGGCCCTC CAGGGTTTGG CTGGTCACCA TGGAGGGGGG 60

GTTCAGGTGC TGAATTTAGG GACCCCAGCA TCTCACAGGT TTCCCCTTCC ATCTTTCCCA 120

GTGGCACTGT GTCTGAGCAG GTGTGCCCAG GTGAGGTTGT ATCCACTGTG TCTGAGCAGG 180

TGTGCCCAGG TGAGGTTGTA TCCACTGTGT GTGAGCAGGT GTGGCTGTTG CAGGTGGAAG 240

TGGGGATATN TGGGCACCTG GGTGCCATT 269

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

TTTCAGGATT TTATTTAAAA TTTATTGTAA TGGGGTCCGC GCAAAAGGAA GGGGTGGAGG 60

GTGGGGTACA TGCAGGGGAC ACAGGAACAN GATCCACATG GCCAGGGNCA CAACTTCTTC 120

TGTCCTGGGG AAGAGGGATG AAAAGACAAG ACCAGGGCTA NGAGCTGGGG TGGAAGAGGC 180

GAGGGGNAAC ACTGGCTGCA TTCCCCNAAC CCCANCANGC ACCTATAGGC CCTGGACCCA 240

TGCCTGACGC TGGGCCCTAG 260

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGGAACTTGT GAGTGGGGAC CCATGATGTA TGGGTCTCAC CTGACTTGAG GTGAATTTTG 60

GAGTGAAGGG CCCTGAGGTC AGCTCCCAGG TCGGTCGTGC TGGGCCAGGC CTGGTTTTCA 120

CAGGGGCTGA AGGATCCCAG TCCACCTGTG TGCATGTCAG GGCTCGGCCG GGAAGAAGCC 180

AGCAAAGTCC CCCGTGTCCC TTGCTGAGTA TTCTGTCACA GACAAGCCTC CATTAAAGCC 240

ACAGCAGTGC TACCCACCAC ACACACCTTG CTGGCCCGGC CACCACTGCT GGCTTCAGCC 300

CCTTNAGCAG CCCATGGNTT AGCAGACCCT CAGATGTAGG TCAGTGGCCT TANCTGTNTC 360

TATCCATGCT GTTAAACTCC CTGCCTCCAA CTGGGGGTCA CCAGT 405

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 400 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

CCATGATCTT ATTTATTACA TCTAGTTTTT CTTTATACCT CTAAAAAAAA GTGCCTTTTA 60

GATTTACAGC TTGTGCTTCT AAAGCAAAGG TTAAAACATC ATGCCCCAAA GGAAAACAAG 120

GTAAAAAGGA AGCTGCCATA TAAGCTCTTA AAANTTGTAT GTTACAAGGT TCTAAAATCT 180

CTTCAGCACT GGTTGGTTGG TAGATTGTAC GACACTGACA TGGTGCTTGG GAGGGTCATT 240

TATCTGATGG TTGGAGCAGC ACCATGGGAA AGCTGCCCAG ATGGTCTACT GAAGTCCTTG 300

GCTGTGCACA GAATGGGCCC AAGGGCCAGN AATTCATGAG TCCGGGGAAC TTTGGNGGTC 360

CTTACTCAAT CTCCTTAGTG CTAAAGNTTC AGAGTCTCAA 400

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

GTCCTCTGGA TTGCTTCGTT GGTTGCGAAC TTTAAGAATG GCAAACTGTG ATTGGNTCCG 60

ATTAAGACAA GCTTTGTAGT TTTCTTCGTG TAAACACCAA ATCCCGCCTG GGCCATGAGG 120

TAGCAGAAGT GGGCCGCATC CAAGAGGCCC CTTGAACCCA CAGTGTCGCC CATGGTAGCC 180

ATCGTCCTGG ACTCGACGTC CATGTTGTTG TTCAAGTTGG ACAAGACCAT GGCGAGGTGC 240

GGCCTCCAAT CTCCCCATTT CTCGTCTCCA CAGCACGTGG ACGCGGCAGG CATCCGTCCG 300

GACATGAGCT GGTAGACTGT CTTCAGAGGG TCGTTGATTK GGGAGGCTTT TTAGCAAACC 360

TKGGTCATGA CTCGGGCGTG TGTCCGGCTG TTCCATCTTA CTTGCAAGTA GCAGAGCGTG 420

ACCCCACAAG GCCATTCTTA ATT 443

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

ATTGGAAAGG GTTTTAAACG GAGTCGGAAC CTGAGTAGAT TTCCAAATTT TACAGCCAGG 60

ACTACAGAAG TGCATCATTC TAGAATGTGT AGACCTGAGT AGCTTATACA CTACAGAGCA 120

CTTTGCTTAT TTGAAAGTAA TTCAGCAACA GGTCACTTTG GGATATAACC TGAACCTTTT 180

TTTGGAGTGG GGTGGGTAGA CTACAGTAGA CACAAGGGCT GGACATGCAG ATGCTTAGGG 240

GATTAGCGTT TTTCATAATT TGTTCTGTTT GTCAGTTCAT TCCTGTGTGT TCTTACCTCT 300

ACAAAGGTAC ATTACACATT TTARGTTTTT TAGTGACCTT TAACCATGTT ACTTGAAGCA 360

TTTTGGAATA TAAAGCTATT TTA 383

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 241 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

TGGTSTGTGG CTCAGCTGCA GCGGCAGCTA AGTGGGTSTC CAGGGGAGTG GACAAGCAAT 60

TCTCCTGTCA TTTGCAACTT TCTTCAGGAA CTCAGATAAA GAACACTTGG ATAACGATGA 120

TGCCTGTAGA GGGATTTCAT CTGTACCATC ACACATGGAA GAGGAGTTTC TAGGTCAGGA 180

AAGGCAGCTN CTAAGCTAAA GGTTTCTTGG TCCCTTNGTC CTGGCATGCC TTAAGGAGGC 240

G 241

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

CTGTCAGTAA TGGCTCACTA AAGGGCCAGC AGTTTAAATT ACACAGGTTG CACTAAAAGC 60

TGCAGCTTTG GCCAGGCAAG GTGGATCACG CCTATAATCC CAACACTTTG GGAGGCCGAG 120

GCGGGCAAAT CACCTGAGGT CAGGAGTTCA AGACCAGCCT GGCCAATATG GTGAAACCTA 180

AGCCTCTACT AAAATTACAG AAATTAGCCG GTCGTGGTGG CACA 224

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 377 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

GGAGGCTGAG GCGGGCGGAT CACGAGGTTA GGAGATGGAG ACCATCCTGG CTAACACAGT 60

GAAACCCTGT CTGTACTAAA GATACAGAAA ACTGGCCGGG CGTGGTGGTG GGTGCCTGTA 120

GTCCCAGCTA CTTGGGAACT CGGGAGGCTG AGGCAGGAGA ATGACCTGAA CCCGGGAGGC 180

GGAGCTTGCA GTGAGCAGAG ATTGCGCCAT TGCACTCCAG CCTGGGCGAC AGAGTAAGAC 240

TGTCTCCAAA AAAAAAAAAA ATAATAATCA AAGCTCTTGG ATTTATAGTT TGGTCCCCAG 300

CCTTGTTTTG ATCTTTCCTT TATCCTGTTT TATTGCCATT TACCACGTCC TTTTGGAAAC 360

ATCCCTTTCA ACTGCTG 377

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

GCAGCGGCGC CGGGCGAGGA GGCGGCAGGG GCGAGGAGGG GGCGGCGGGT GGCGACCCGC 60

AGGAGGCCAA GCCCCAGGAG GCCGCTGTCG CGCCAGAGAA GCCGCCCGCC AGCGACGAGA 120

CCAAGGCCGC CGAGGAGCCC AGCAAGGTGG AGGAGAAAAA GGCCGAGGAG GCCGTGGCCA 180

GCTCCGCGCT GCTAGGCCCC CTTCGCGCGG GCCCGGCGCG CCCCCGGAGC AAGGAGGCAG 240

CCCCCGCGGA GGAGCCCGCG GNCGCCGCAG ACT 273

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 286 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

TTTTGGTCAA ATAAATCAGA GTACTACAAT CATCAAACAT CTGATTCATT TAACATGTGA 60

GCATCTATAC CTGCCCATTT GTGTGAATAT TCAGTATATA TCTCATACCT ATTCTCATGC 120

CTTCATTTAT TGTGGTTATG GCTGTAGATA TGGAAAAAAC AGTAGCTGAG ACATTTTTAT 180

TATGAACTAT ATTATACCTT AATCAATCAG TCAGAAAATG CTTAGGAAGA AGAAATGCAT 240

GATTGTAAAT GCATGATTTC AACATGCTAC CCGGCCAACA AAGTTG 286

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TGCCCAAGGA AGACAGAACA TGGAGAACCG TCAAGGCAGG AACCCCACAG ACTGTCCCTT 60

CCAGCCCACA CTCTGCCACC TCCTGGCCCT GTCCCAATTC TGAGCCAAGG CCTCCCCGAG 120

CCAGAAGTTG CCTGGTCCTC TGTCCCCACA GTGACCTGAC TGGGGGTGAG GGAGAAGGAG 180

GAGAGAGCCC ATGTGTGGTG TGTGTGCCCC TGAGAACTTC GTGGTGACTG CCTTTGGGAG 240

CCCGCAAGTG CCCAGAGGCA GGGGTAGCTG AGTTCCTGGG AGACCCCTTT TTTTCCCCCA 300

TGTTCCCCAG AGGGCAACGC CATCAGTAGC AGTGTGGTGT TT 342

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

ATTACCCGGG CCCCGCCTCC CTAAAACAGA TCTACGGACC TTAACCGACG CCATGCTGAG 60

GCTCATTCCA TCCCTGCRGA CGTATGCAGA GCCGCTCACT GCTGCCATGG TGGAGTTCTA 120

CACCATGTTA GGAGGAATTC ACCCACGATA CACAACCTCA CTATATCTAT TCACCCCGTC 180

AAATGACTAC GTCGCTGAGA GGCATCTTTC AAGCGCTGAG ACCTCTCGAG ACCCTCCCTC 240

TTGAAGCGCT CCTTCGCATT TGCGCACATC AAGCTCTGCC TCTCTTCCCA GATAGACTCC 300

TARGGGATGA GGAGACCCCT TCGGACTGAA TGACAACATC GACACGGTTC CTCTTCAAGC 360

CACTTTCCCT AACCTTCGCC AGAGACGAGG GC 392

(2) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

GTTATAATTA TCTTGTTTTA TTATTTATTG TTTATCTCTT ACTGTGTATA ATGTAGAAAT 60

TAAACTTTAC CATAGGTATA TACATATTGG AAAAAGCATC TTATATACAG GGTTTGTTAC 120

TATCTGTGGT TTCAGGCATC CACTGGGGGT CTTGGAACAT ATCCCTTGCA GATAAGAGGG 180

AACTGCTGTA TCCATAGAAT AAAAACACCC CATCTTGAAG ATAGGAGGTT CTGTAAATTG 240

GGATGGGGTC AGGGAATCTG AATTTTAAAA GTTTCCCATG TGATTTGATG CCCAGCCAAG 300

GGCTGGGGAC CACTGTCTTG AAATATAATG CTGAGGAAGA TACTGTCTTT GGATTTTCCT 360

GGTAATTCCG AGTGCAAATT CTCAGGCTGG AACCTTATGG GCCTTG 406

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 453 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

GAAAACTTTG CCATGGGTCA GTTTTATTGG AAGTTCATTT TCCTGAATGT TTGGAAGAAA 60

GTCTAGTGAC TCAGGATAGC ATTTCTAATT TCACAGAGTT ATTTTTCCGT TATGAAACAC 120

AGATTGCCTT TGAGGTCTCC TGTTTCTACT ACTGCCCCTC ACTTTTATGT GGGCCTCCTC 180

TTTCCTTTGT TTCTGGAGAA CCTTTTCCTG TTCAATTCTG TTTTAATTTT CAGCAGTTTT 240

TTTTCTGTGT GAGTGAGGCT GTTTCCTAGC AGGGAGGTCT GGTTGGTCAT TTTCAAGTTC 300

ATCAGGGCTT CATCAGGGCT TGTCCACTTC AACCCTTACG CTATAGGNCC CTNTGCACCA 360

TCTGCANTCT TCAAAATGTG CCCACTGGTT CGTTCCCATG GANGGCTTGT TGGTAATTTG 420

CGCTTTTAGG GGGGGCCATG GAAGGAGCAA ATC 453

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

TTTACTCTTA AAACTGTTAC AACAGAATCA TGGACTGACA CAGGTAATGG CTGAGCCATA 60

AGCAAATCGA GAAGTACAGA AATGTCCCAC CCCAAACAGC TGCGGAGTAC ACATCACACA 120

GGGCCTCTGG TCCCGGCCTT CTCAGGTGCT CTGGAGTGGA GGATCCTTTG AGGGAACTCT 180

GACCACTCCT GTTGTCTACC TAGAGAGCAC GCCACTTGGG CCACCTACCC CCAACCTTTG 240

GCCAAAGGAG TGAAAGGACC TGGAACCTGT CGTCAACCTC AGCAT 285

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

CTAGAGGGCA CTCTGTATAC CCGTCAGCTC CTGGAGCCAT TCATTCTATG CTGGGCAGAC 60

AGGCTGTGAG AGGACATGGG GGACGGTGGA AAGGNTCCAA AGACGAAGCT GTNGTTTATC 120

CTTGTTGGTT TTACACAGGG AATGATGAAA CATTGAAGGG GTTTAATAAG CTTTTCCTAA 180

AACATTTTCC CCCTAAACAG GCTGGCACTA TGTCGAAGCT GCCCAAATTT GAGATTGATT 240

TACCAGCTGC GNCTAAGTCA ACTAAACCCA NGCCTTTCCG AAAGAGACAT CGCAANTGGC 300

TTACCCAANG TANTGTCCCG TTTTCAG 327

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

GGTGCTATGG AGAGCCGGCC GTCCTCCAGG GGTGAGCTGG GGAGGCTTCT GCGGTTCTGG 60

AGTCCCGGCG ATGGCGCCAG TTCCCCAGCA AACCCCCTCC AGAGCTGCCC CCGGATGCAC 120

AGACAAGGAG GGGGCTTGGG AGTGACTTGA GGCTGTGACG GGRTCGCCCT CGGTGTGGGC 180

AAGTGAGTCC TTTGTGGCCA AGAGGTCAGA GTGGTCCCTG AGGCTGAGTG GAACACAGAC 240

CCGTGGCCCT CATAAAATTA AACATAAAAC CACAAAAATG GGCGCAACCA GACAGCATTG 300

GCTTTCAGAC AGGCAGGGAC ACGGGGGCCC CTTCGTGTTG ACCTGT 346

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

TTGACCTCAA CTTACTGACC AATCCCGTAG GTATCGAATA GAAGCATTTG TTGCACTCTT 60

TTTGTGAGCC AGGCCCTGTA GGAGGGATTG TGGATGGCAA AACCTCAGGT TCTGCCCAAA 120

TCCTCCCCTT GGGGGCTGGA GGGTCTCTAG TTAATTGGCA TTCCGGTGCT TAAGGCCACT 180

TTTGGGTAGA GGTTTGGCAA GGATGGAGTG TCCAGACCTA TGATCCTCTA AGAACTTTAC 240

CTTTTAAAAA CAGCCACCCA AATGGTGGTG GCGTGGGGAG CAGGTGGTGG TGAAGGGACT 300

GGGGGTGTCT GGCCATKGCC ACGTACCAGA GGAGACTCTG TGAGCCCTCT CCCTGCCTGA 360

GGGACACTTA ACTTTTATAG CACTACATAG GGTCAACG 398

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 321 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

AGACAGCATC TGGCTCTGTC ACCCAGGCTG GAGTGCAGTG GCGCAATCTC GGTTCACTGC 60

AACCTCTGCC TTCCAGGTTC AAGTGATTCT CCTGCCTCAG CCTCCCAAAT AGCTGGGATT 120

ACAGGCATGT GCCACCATAC CCAGCTAATT TTTGTATTTT CAGCAGAGAC GGGGTTTCAC 180

CATGTTGGCC AGACTGGTCT CGAACTTCTG ACCTCAAATG ATCTGCCCAT CTAGGCCTCC 240

AAAAGTGCTG GGATTATAGG TGTGAGCCAC TGCGCCTGGC CCTTGGGTAA ACACTTCAAA 300

TGCAMCCAAC CATTAAAGGT A 321

(2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

GAAACTTATA GTCTTGCCTC CCAACCTTCT GAACACTCCA GTAGAAAAAT CTTCTCGCCT 60

ACCTTTATCA CCCCACGACC TACTAGCATT TCTTACTCTC AAAAAAAAATC TTTTCTGAAA 120

AATCAAGACA GAGTGCAAAC AATCAGCATA ATTTTATTAT GACARAACTT TTAAATTTTA 180

TCCCCCTCTC TGAGAGKTCT GCTAGGACTC CTTCAGATAA GTGAAAAAGA AAKTTTTTAA 240

AATTTATTCT CAAATCCGAA TTCCAATCTG TATAAAAAGG GCGATTCTCC CTC 293

(2) INFORMATION FOR SEQ ID NO:173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

GCTTGGTCCC GTTCCTCAGG AAAAGGATGG ACCTTCTCTT CTTCTCAGAT GGTCCCTTCC 60

ATTCCCCTGA AACCTGCATG AGAGCTCCTA ACATGTTTCT CCAATGCAAT CAAGCCTAGA 120

CTCCAAATGT CCTCCCAGCT CACCTCCATC TATGCATCTC ATCTCTGGAT TTGGTGATCA 180

GACTCTATAT TGACAGTAGG ATCTCAAACC CTGCATCCAT CCTTCCTCCA GCAAGCCCTG 240

CTAGCCACAT GAGGAACAAG TTTCCGTGTC TTCATGACTT CC 282

(2) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 353 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

CAAGAGGTGG GAGAGGTAGG GGGCAACTAC AGCTCCCCAC CAGCCCCACC AGGGGGAATG 60

GACCCCTCCC TGCCTCCTGC CCAAGTGGCT CCCCCTGTAT TATGGGGGGG ACTTTGTGCA 120

AACTCTGCCC CGAGGGGGTG GGGAGGGTGG AGGGTGAGTG TGAAATGGCA GCGGTTGGGG 180

CTGGCAGCTG TGCTACTGGG CACTGGGGGG CTTGTAGGGC TCCAGGAGGA GGGCCGAGAA 240

GGTGTTGACC TTGTCTGCCC CCCGCACCTC ATGGGGTAAC AGCGGCAMTT TCACGATGTG 300

GAAGTTCTTC ATACAGGTCC TCCAATCTGG TCCAGATACT TGGCCTGGGT TCT 353

(2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GCCCATGCCC TTGTGTACAT AATCTCTAAT ATTTATATAT ATTGATATAG AATTCTCTCT 60

ATAATATATG TCATAGAATC TCTCTTGGGC CTGGCCGTGG AATGTCACAT TAAGAAAACA 120

TGCTAAGACT GGCCAGAAAA ATGGATATTT CCCAGACCTG GAGGATGGTG TGTGGGATGT 180

ATAGGTGAGG TCGTGGAGAA GATAATAAAC TCATTCCCCA AGATACCCTC TTCAACACAA 240

GGACAACAAC GAACGTGTCT GGTGGGGAAG GGGACAATGG AGGGCGAGGA GTGGAAGATT 300

TGGATTTTCA TTTAATAAAG TCAATTGAAA AATGAAAGTG CACCCCCCCT CCAAAAAACA 360

GGAGATTCAT TTAGCAAGAG CCGTTTTCATT CACA 394

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

ATTGGGACGG GCCCCCCTCT GAGGCGACGG ATCGATAAGC TTGATATCGA ATTCCTTGAT 60

NTTTTCTAGT GTTATGGTTT TCTCCCACTC CAATAACTWT TCATACCTKT GGTCTKAGTT 120

TTTCCATCTA TAAAATCATG TGCTAAATAA TTAACTATCA TCTCTATCAT TGTCAGACTA 180

CACAAAGCTT CCAGCCTGGG CAACAGGAAC CCTGTCTCTA AAAAAAATAC AAACATTAGC 240

CAGGTGTGGT GGTATGCGCC TGTATTCCCA GCTACTTGGG AGGCTGAGGT GGTAGGACTA 300

CTTGGGCTTT AGAGGTCAAG GCTGCAAGTG AGCTGTGATT GCGCCACTGC ACTCCAGCCT 360

GGGCAACAGG GCAAGACCCT G 381

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

GATTTTATTC AAACACAGGC AAGAACAATG ACCTTCAGAG CTGGGTAAAA ATAATAAGTT 60

AAAAGCATGG TTAGAATTTT AGACAATCAG ATAAAAAGTT TGAAGGAAGT GATTTCCCCT 120

TCCTCTCCTA ATTGATTAAT TCAACACAGC ATAAAAATAA TTTGTATCTA TAAAATATCC 180

TTGTTCCCAC ACAAATGAAC TGGAGGTGGC CCTAGGATTT CCTTGACTAT GCACAATGCA 240

CACAATCTAC ATGTCCCTCC TCCCCAACTT TTAAGGCAAA AATGGTCCTG CATCTTCAGG 300

CAGAGGGTGG GCTCATGCCA GCAGTCAGCT GTGGTCAAGG ACACTGGGGG TGCGTTTYCT 360

CCACCGAAAG ATGCCTGCTT TGGGTCCACT TTGGGCGCGG GATCCCATTT TATTTTCTAG 420

CCTGTGCCTC ACCACAGGGA AAA 443

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

TGGGGGACCA GCATTGCTCC CAGCTGAGGG CGCCGTCTTC CTCACCACGT ACCGGGTCAT 60

CTTCACGGGG ATGCCCACGG ACCCCCTGGT TGGGGAGCAG GTGGTGGTCC GCTCCTTCCC 120

GGTGGCTGCG CTGACCAAGG AGAAGCGCAT CAMCNTCCAG ACCCCTGTGG ACCAGCTCTT 180

GCAGGACGGG CTCCAGCTGC GCTCCTGCAC ATTCCAGCTG CTGAAAATGG CCTTTGACGA 240

GGAGGTGGGG TCTTACAGCG CCGAGCTCTT TCCGTAAGCA GCTGCATAAG CTGCGGNTAC 300

CCGCCGGACA ATCATGGCCA ACTTT 325

(2) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

GAGCATGCCC CCGGAGTCCC CAAGATCCTG GTGGGGAACC GCCTGCACCT GGCGTTCAAG 60

CGGCAGGTGC CCACGGAGCA GGCCCAGGCC TACGCCGAGC GCCTGGNCGT GACCTTTTTT 120

TAGGTCAGCC CTCTTTGCAA TTTCAACATC ACAGAGTCGT TCACGGAGCT GGCCAGGTTC 180

GTNCTGCTGC GGCATGGGAT GGACCGGCTC TTG 213

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

AGCTTTATCA CATTATACAC AAACATAGAA AACAGTGTTT CAGAAGAGAA GCAAAGGCCA 60

TTGGCTTCAA ATATTTATGC AACAATGAAA ATGTTCTCAG CCCTTAAATG AGCACTTGTG 120

ACTTGTCCAA CAGTGAGATA ACTAGTCAAT GGAAGAGTTC AACACTAGAG CATGTATCTC 180

AGTCTGTTCT CATATTGCTA TAAAGGGCTS CCTCAGACT 219

(2) INFORMATION FOR SEQ ID NO:182:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

GTGTTAGTCT GTTACCCAGG CTGGAATGCA GTGTTGTTAT CATACCTCAT TCCAACCTCT 60

GGCCTCTAGG CTCAAGTCAT CCTCCCACCT CAGCCTCCCC ACTAGGCTGG ACTACAGGTA 120

CATGCCACCA TGCCCAGCTA ATTTTTGTAT TTTTTTTAGA CAGGGGTTTC TCCATGTTGG 180

ACTAGGCTGG TCTTGAACTC GTGAGCTCAA GTGATCTGCC TGCCTCGGCC TCCCAAAGTG 240

CTGGGATTAC AAGCGTGAGT CATGGTGCCT GGCCTAGTTT GCTCTTATTT TTTTTCCATC 300

TTTGCAGTTT CTAGGCCACT GGGAACAGGC TGCAGAGCTC AGAGTCCACA GCTGTGAGGC 360

TCCATGTTGC ACCATCAAAA AATAAGGTGA CGAGAGTCCT GGGTTTCCCA GTGTCACGGC 420

AAGAGGGGTT ACTGCTCACG GGTACACACA G 451

(2) INFORMATION FOR SEQ ID NO:183:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CCAAGTTGAC CCGCCGAACC ACCGACAGGA AGAGTGAGTT CCTGAAAACT CTGAAGGATG 60

ACCGGAATGG AGACTTCTCA GAGAATAGAG ACTGTGACAA GCTGGAAGAT TTGGAGGACA 120

ACAGCACACC TGAACCAAAG GAAAATGGGG AGGAAGGCTG TCATCAAAAT GGTCTTGCCC 180

TCCCTGTAGT GGAAGAAGGG GAGGTTCTCT CACACTCTCT AGAAGCAGAG CACAGGTTAT 240

TGAAAGCTAT GGGTTGGCAG GAATATCCTG AAAATGATGA GAATTGCCTT CCCCTCACAG 300

AGGATGAGCT CAAAGAGTTC CACATGAAGA CAGAGCAGCT GAGAAGAAAT GGCTTTGGGA 360

AGAATGGCTT CTTGCAGAGC CGCAGTTCCA GTCTGTTCTC CCCTTGGAGA GCACTTGCAA 420

GCAGAGTTTG AGGCTCAGCA CCGA 444

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

GGCAGAAAGA GGAAGGAGAC AGTGCCAGGA GGAAGAAGGA AGGAGTCCCT TAGCTCTCTT 60

CATTGTCCCC TTTACTTCCT GCTATCTTCT TCTCCTCTTC TTCTCTCTCT TGCCTNTATG 120

CCTGTATTTC TGGCAATATG ACAGGCCTGC CTACCCAAGA TCAGAACTCC AAAACCACTC 180

CCACCCCTGA AGGTCGGGAG GGTCTTAGCA GCCCTGGGTG GCTGCCTGTG CTCAGGTCCT 240

CAGCTCCATG GGAAATAAAA ATGGCACCCT GAATCTCTAG GATTTTGTCA CTTTGGAGTC 300

ACAGCAAAGT TCTCTTCCTC TTGTCCCCCC GTTTGCTGCT CCTTGGGTTA TAGGACATGG 360

TAAATATTTA TTACTTTCAG GGAACCAGTA TTTTATTAG 399

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

CAGAGACACT GGCCCAGCTA TTTTCAGCAG GGACAGAGTC GAGGCTCACT GGGGATGGCT 60

TCAGAGGACA CTGAGGCCCC TCTCAGGGAG GGCAAGGCAC AGATACCCCA AATTCCACCC 120

CACGTCCCAA AGGTCTCCCA GCGGGGCTGT CCAGTCCATG TCAGCAGAAG GCTCTTGGGC 180

GTGTGAGGGA GGGTCTTGGA GAACTAAGCG AAGGAGGCAA ACGCCAGGGC CCCTTGCAGG 240

TCAAGGCACC ATGTGCACCA CTT 263

(2) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

GTTCCAATAG CTGGTTTTAT TCTCAGCACA AAAGGGCCCT GTGTAAAAAC CAGAAGGATT 60

TTGTAAAATA TCAAAATGAA TATTTGGCCT GGAGGTTGGA AAGTGAAGCA AGGCTGGACA 120

TAGAAAAAAA CTGATCAGTA GTTATTCAGG ATATTATTTA GGATAAATGA AATAGGAACT 180

TAGGGCCATC TCTTACTTTT CTACAGGTTC TTATCTGGGT CAATGAAGAA ATTGTGTTTA 240

TCTTGCTGCC CTTGCATCAG GTTTTTTGCA CTAATGGAAA AAAGCCGGGC GAAAAACAAA 300

ACCCAATCCT TTCAGTCCTA GCTTTTACAT CTTCCCCTTG CAA 343

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 209 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

GGTGCGGGTC CACCCCTTCC ACGTCATCCG CATCAACAAG ATGTTGTCCT GTGCTGGGGC 60

TGACAGGCTN CAAACAGGCA TCCGAGGTGC CTTTCGAAAG CCCCAGGGCA CTGTCCCCAG 120

GTTTCACATT GGCCAACTTA TCATGTCCAT CCCCACCAAG CTCCAGAACA ACCACCATGT 180

CATTCACCCC CTGCCCCAGG CCAACTTCAA GTTTCCTCCC CCCCACAAG 229

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 284 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

CCAGCAACTC AAATTCACCA CCTCGGACTC CTGCGACCGC ATCAAAGACG AATTTCAGCT 60

ACTGCAAGCT CAGTACCACA GCCTCAAGCT CGAWTGTNAC AAGTTGGCCA GTGAGAAGTC 120

AGAGATGCAG CKTCACTATK TGATGTACTA CGAGAKGTCC TACGGCTTGA CCATCGAGAT 180

GCACAAACAG GCTGAGACCG TCAAAAGGCT GACGGGATTT GTGCCCAGGT CCTGCCCTAC 240

CTTTCCCAAG GAGCACCAGC AGCAGGTTTT TGGGGGCCAT TGAG 284

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

GGAAGGATGA GAAACAGATT TCTGCTCACT TCATGGGCTG RCCTRGRATT GACGATGGTR 60

CAAACCCAAG ATTATCCTCA TGTAATTTAT GAAGATTATG GAACTGCAGC GCATGACATC 120

GGGGACACCA CGAACAGAAG TAATGCAATC CCTTCCACAG ACGTCACTGA TACAACCGGT 180

CGGGCACATC TCKCGGCCTA TGCTGCCGGT GGTGC 215

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 153 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

TTTCATATGG AAAGAGCTAG TACAATCACA TATTTGAAAG GAGAAACAAT AGGTACTGAA 60

CCGGAGGGAA AGGGCGAGGG TGAGTGTGCC AGCACCGGCC TGGTGAATCC ACGATTCGGT 120

TTCCCATCCA AGGGTAAGTT TCCCAAAATA CCG 153

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

GTATTTATAC ATTTATTTAT ATATGTATAT TTACTTCAGA NGAAACGAAC ATTTCGGGGA 60

CAGGAAGCAA GCAGGCCCGG GGCTGCTTCC CTCACTGCCC ACCTCAGAGT CAGAGTTGGC 120

ACATGACAAA TACCAAGCTC AGGGTGAAGA ACTGGGAGTT AACTGGGAAG TAGGGKGCGC 180

TCTATGCACA CGCAGGCTTC TAAGGGTGCA CGGTATGGGC AGNKGGTTTG CACTGGGAGG 240

CCCTATGTAC AGCTTGAAAG CTAGGGGTGA GATTAGCCCA GTGACTACAG GAACATACGT 300

CAAAGTTGAG AGAAGA 316

(2) INFORMATION FOR SEQ ID NO:192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

GTGGTTTTTG GTTATATGCA GCTTTTGACT AGCATGTATT GTGTCTTTTT CTCCTCTATG 60

AATAATTTTA TATTTCATGC TACTTCTTGA AAGTTTACTC TTTGATGCTC TAAGAGAACA 120

GCCAGATGGT TTATATGAAT AANCTTTATC TGCAGGATGG TGGATTGGTA AATNAGGAGA 180

ATGTTGTTTG AGATATCAAG ATTTATGTCT GGGAACTAAA ATATATAATG CCAAATGTGT 240

TTTTGTCAAT TACTAGAGAA TTCTGTGCAA ACATATCATC TCTTCACATG CTGCACACTT 300

TGCTTTTTGT TAAACAGCAG GTAGTAGACA GACCAATACC AGTTTCGCGT TAAGGCTTTT 360

(2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 397 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

GAAAAGACCA AGGAGATGGT GAAGACAGCA GAAGCCCAGA AGCAGCAACT GAAGGAGGAG 60

CAGGGGAGGT CAGCAAGGAA CGGGAGAGTG GGGATGGAGA GGCTGAGGGA GACCAGAGNA 120

GTGGAGGGTA CTATTTAGAA GAGGACACCC TCTCTGAAGG TTCAGGTGTA GGGTCCCTGC 180

AGGTTGAGTG TGCCAAAGAG GGCAATCCTG ACTCTTCTGA GATGGAAGAG GTAGCCCCAC 240

AGCCACCTCA GCCAGAGGAG ATGGAGCCTG AGGGCCAGCC CAGTCCAGAC GGCTGTCTAT 300

GCCCCTNTTT TCTTGCCCTG GGTTGCCCTG GGGCATGCGT GTAGTTTTCA GTCTGGTTCA 360

GGTGCAACAG GGTCCGTTCT GTGCCTTTGG TGCCCCC 397

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

GATTATTGGC TTTGCTTTCA TAACATGTAT TTTTAAGTAT TTACTCTCTT AATGGCCCTC 60

GTGTCTATTT TATACATCAT ATCTCTTAAT TCTCTAGATG GAACACTGAA GGACAGGAAT 120

TAAGTAAGTG ACTGGCCATG CAAGGGTTGG AAATTTTACT GTATCCCTTC CTCRGTAGAA 180

GTTATGTTAA ACATTCAAGC AACCACATAT CTAACAGAGG AGTTT 225

(2) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

ATTACTAGAT ATTTGTATGT TAAATTATGT GGGTTTTCAA ATTTGTGGAG AATAAGTAAT 60

AGTGACATTA GTTTAAGGAC AGTGTTTCAT CAGGGCATTA TTTTAATGAA TCTTATATTT 120

AAATGTCTGT TTCAGGAATT CATGTGAATC TTTCTTTTTA TAGAGGACCC ACAGGCATGA 180

NTTATTTACT CCTCCGGTGA TAGGTTCTCA CCCTGATGAA AGCGGAAGCA AATTCCAGGT 240

TAGAACATTA TNCTAGTTAT GTAGGGGGGT ATAAAGTGTG TAAGTTTAAT ATTT 294

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

TTATTTTTCT CTAAATTTTA AAATAGAAGA CTTTAATGGA AAACATTTAG TACCATCATG 60

TCAMCCTGAA TGCCAGCAAT ACCTCGACTT TTACACACGC AGGAAGCCTA GTAAAAGCCC 120

CGTCAGTAGT ACACATTTCT CTATGGTCCT TCAACAGTTT TTCATATACA AAATTTTCTG 180

CTATTTTTGC TTTTGCAAAC AGCAATAACT TTTGGGTTTC CCATATGACC ACC 233

(2) INFORMATION FOR SEQ ID NO:197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

AAGATATCTA CCTGGAGTAG CTGTGCAGCC CCGCCCTCTG CTTCCCCCAG CCCTCAGGCC 60

AGTGCCAGGA CAGCTGGCTG CTGACAGGAT GTGGCACTGC TTGAGGAGGG GCACCTGCCA 120

CCGCCAGAGG ACAAGGAAGT GGGGGCCGCT GGCCAGGGTA GGGAAGGNTG GGGCAATGGG 180

GAGAGGCAAA TGCAGTTTAT TGTAATATAT GGAATTAGAT TCATCTATGG 230

(2) INFORMATION FOR SEQ ID NO:198:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

TTCTCCTGGG GAAAGGGCTG TTGCTGAAGT GGCCGGTTTT TTTAAGCATC GACATTTGCA 60

TCCAAAGGTT CAAGCAGCCG CCTCAGGTTC CARAGGCTTC CACCTGATGG CTGCACTT 118

(2) INFORMATION FOR SEQ ID NO:199:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 268 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

TAAATGATGG AGTTAAATGA TGTTGTCAGT GCCTATTTAA AAAACTACTC TTCCCCTTCT 60

CTATGAGTTC TACTTTGGTA AATATTAATA TTTAACCAGT TAGTAAAACT AACACCACTA 120

TTTCAATTCT CTTTTGTGCA TAGTAAGTAA ATTTTGCTTT ACTTACTTTA TAAAAAAATA 180

GTTTACATTT TATAAAGCAG GTTTTAGAAA AACGGTTTAC AAGAAAGTTT GCCTCCATTT 240

CACTGCCAAT TTAAGCACAG GGGAAAAT 268

(2) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 422 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

GGAGTGAGTT TGTGAAAAGC AACAGGGGTA NGACAGGTTC AACGAAGGAC ACAGACAGTG 60

CCCTGTTTTA GGTTCCAAAT TTGTTGTTTT TAATGGGTGG TGGGAGCTGA GGAATGATGT 120

CATTTGGAAG GGGGAATGAT TTGTGAATNA TGCACAATAT GTAGGCATGA TGGAGAAGAA 180

TGTGCATCGG TCTCTTGGGA TGAAAACTGA TGTGTGTGAT AGGAGTATCC CTTTGGAGCC 240

AAAGGTGGTG AAAGCCCTGC TTCTGGACAG TCCGGCTCCA ATCTGTATAC TGTTTGTCTG 300

GGATGCTGTA CTCAAATACC TGCTGGTCCG AATGAGCGAT GACAAGGTTG TTTGGTATTG 360

GGGGCAATAG CCATAGCAGT CACTTGGGAA ATTGTAAGCA GGCACCGTGC AGTGAAGTTT 420

TA 422

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

ACTCCACGCT GATGAACCCG ACGTCCATTT CTCCAAGAAA TTCCTGAACG TCTTCATGAG 60

TGGCCGCTCC CGCTCCTCCA GTGCTGAGTC CTTCGGGCTG TTCTCCTGCA TCATCAACGG 120

GGAGGAGCAG GAGCAGACCC ACCGGGCCAT ATTCAGGTTT GTGCCTCGAC ACGAAGACGA 180

ACTTTGAGCT GGAAGTGGAT GACCCTCTGC TAGTGGAGTC CAGGCCCCCA GACTACTTGT 240

TACGAGGGCT ACAACATGTG CACTGGGTGC CCG 273

(2) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

GGACTCCAAC CCCCCAGGAG GCCGAATGCT GAGCTTGGCA ATGGTGGCCT GGATGGAGCT 60

GATGGGCACA TCCCCACCGA GGACCAGGTC CTGGGAGTCC TGAGGAAGGT GGTTCTTCTG 120

GCTGATGCTT GCACTGGCCA AGGGTTTGCA TGGAGGAGGC ACACCATGGC GCTGCAGGAC 180

CTGCTCCACG TGTCTCACCA CTGCCTCATA GCAGAACCTG AGGTGCAGCT TCTCCTGCAG 240

CATGTGCTTT CTCTGCTGCC GCATGCGCCG CACCAGCTGA GGCAGCTCAG GGATTCCKTT 300

CCCAGCCTCC ACCTCCTGCA CAGCTGCATA GAGCAGTGCA AAGGCTCCCG TGCGGCCCAC 360

ACCAGAGCTG CAGTGCACAA TGATGGGCGT TTGCAGGGGC CGTGATGCAA GGTAATTTGC 420

GTGCACCTCC TGGGTT 436

(2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 336 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTGCATGTNT TGGGGACACT TACGCCAAGG CGCCGCGTTC TCATTAGGAG CTGGGACCAG 60

AAGTGAATAA GCCAGGTTCC TGTCTCAGGG AGCTCCATAG CAGGACTCAG AACCACACAC 120

GGCCCTCTAG GCATTTNTGA AGCTCTGTGC TTCATTTTTT TTGCTTTGCC TCTAGTTTTG 180

CCTTTGCAGT ACCAATGCAG CCAGCCCATG TKTCCCCTCT ATGTGGAATG TTAACGATAT 240

TCCCACTGTT TCTGGTGTCC TTTCTGTAAT CAGAGCTGCC GTGACCATTC CAGTTCAGGC 300

ATCCTGGTGG CCTGGCTTTC TCTGGGGCAT AGAGCT 336

(2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

GGAATCAGAT GCTCAGGTGT CCAAGCAGGG ATAAGGACAG GCAAAATAAA TAACCCCCCA 60

ACCCCCATCG TCACTCTGCT GCAACACGAC ACAAAGGTTT AAAGATCTGG GCCCAAAGAC 120

TCTGGGTCCC TTCAAGCAAG CTCAGGTGGA AGGAGGTTTC CCCACCCCCC ACCAGGCCTG 180

TTTGCCCCAG GTTGCCCTAG GATGGAGGCA GTTCAGACCC TGGGTCACTG AAGCTGATAG 240

GAAGAACTNC GATATCAATG GCCTAAGCCT GCTGTNTGCC CAAGGGAGCC AAGGGCAAGA 300

GCCAAACGGC CAATTTAAAG GACGTGGACC TGGGGGGCCA GAGGAGGCAC CACAGCCGAG 360

GGGAGCCACG CCCTGGGCCC GCAGGGCACA TGG 393

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

GAGGAAGAGG ATGAGCTGAG TGAGCTGCCA CCGCTGGAGG ACATGGGACA ACCCCCCCCC 60

GAGGAGGCTG AGCAGCCTGG GGCCCTGCCC CGAGAGTTGC TTGCTGCCAT GGAGCCCGAG 120

CCCCCTCCAG CCCCCGCCCC AGAAGAGTGG CTGGACATTC TGGGGAAGGG GCTGTTGAGG 180

AAGAAGACCC TGCTCCCAGG CCCGCCAGCT TCGACCCCCC CGGTGAAGCC CCAGCTGCTC 240

ACCGTACCTC TNCAGAGCTC GCTGGAGAAT GGCACACCGG TGCAGGAGGA CCCCGAGCTG 300

GTGTTCACTC TGGGTGACTG TNACGTCATC CAGGCCCTGG TTCTCAGTGT CCCACTCATG 360

GACGTNGGGG AGACGGCCAT GGTCACTTCT 390

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 172 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

CTTTACTGTG GGTGTGGGTG TCACTGTCAC TGCCACAGCC ACTNGGAGGG ACACACAGCT 60

TTAACCCCTR TTTGCTTAGG NGAAGGGTGG GGGCATTCAG GGTTATAAAA CTAACTATAT 120

ACACAGAAGG TCCTAGGKAG AAAGCCACCC TGAGCACACA TGTCTAGGCA CA 172

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

AAGGCAATTA GAAGATTTAT TGAATATTGG TTAAAAGTAG ATTGACAATG ACATTAAAGA 60

ATAAAGTGTA ATTTATTTGG TGCTACTTTG TGAATGCTTC CAAGTACAAA TCATCTCACA 120

ATACCATATA CAACATACTT TCAATCACAA CTCAAATATA AAATAACCTA CAAAATCACA 180

TTGCTATAAT CAATATACAA TAATTGTATT TTTAA 215

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

GGAGTTCTCT TGTCCACGGA GAGCAGTGTT GCAGTGTATG GAATGCTAAA TCTTACCCCA 60

AAGGGCAAGC AGGCTCCAGG TGGCCATGAG CTGAGTTGTG ACTTCTGGGA ACTAATTGGG 120

TTGGCCCCTG CTGGAGGAGC TGACAACCTG ATCAATGAGG AGTCTGACGT TGATGTCCAG 180

CTCAACAACA GACACATGAT GATCCGAGGA GAAAACATGT CCAAAATCCT AAAAGCACGA 240

TCCATGGTCA CCAGGTGCTT TAGAGATCAC TTCTTTGATA GGGGGTACTA TGAAGTTACT 300

CCTCCAACAT TAGTGCAAAC ACAAGTAGAA GGTGGGTGCC ACACTCTTCA AGCTTTGACT 360

ATTTTGGGGG AAGAGGCATT TTGACTCAAT CCTCTCAGTT GTACTTGAGA CCTTCCTCCC 420

AGCCTGGGAG ATGTTTTTTG TATT 444

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

GCAGATCACT TGAGGTCAGG AGTTCGAGAT CAGCCTATAT ATGCAAGTAC ACACACAGGC 60

ACTCGCACGC ATGCATGCTC ATGCAACACA CATGTACACT CTACATGTAC AGCTCACATA 120

TGCATCCATA CACATGTGCA TGCTCACCCA TACACCAGCC ACACACAAGT ACTCATACGC 180

ATACATGGCC ACACACAAAG TACACACACG TACACCATAT GCATATGTAT GCACTCATAC 240

ACTCATACAT ATGTGCCCCC TCAGAGAAGT ACACAAGTGC ATGCGCATCA CACATGCATA 300

CGTGCTCATG CATACACACG GGACATTTCA TACACACG 338

(2) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 371 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

GAGGAAGTAG AGCCTNAGGA GGCTGAAGAA GGCATCTCTG AGCAACCCTG CCCAGCTTGA 60

CACAGAGGTG GTGGAAGACT CCTTGAGGCA AGCGTAAAAG TCAGCATGCT GCAAGGGGAC 120

TGTAGATTTA ATGATGCGTT TTCAAGGGTA CACACCAAAA CAATATGTCA ACTTCCCTTT 180

GGCCTGCAGT TTGTACCAAA TCCTTAATTT TTCCTGAATG AGCAAGCTTC TCTTAAAAGA 240

TGCTCTCTAG TCATTTTGGG TCTCATGGCA GTAAGCCTCA TGTTATACTA AGGGGAGTC 300

TTCCAGGTGT GACAATCAGG TTATTGGAAA AACAAAACGT GGTTTTGGGA TCTGTTTGGG 360

AGACTGGGGA T 371

(2) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 295 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

CCTCCCAACG TGTTGACATT ACAGGCGTGA GTA [illegible] TAGCATAATG 47

TTTTGCATAG TTGTCAGCAG ATAAATATTG AATGACAAAA CTCAGATGGA GGAAAAAGAA 120

CAAAATAACC TAGTTCTCAG AAAGATTTAA TGAGCAAATG GGAAAATGTC AAAAAGATTT 180

ACAGACAGGG GCATCTTAGA GTCACTGGAA TCACACAGGC CTTCCCTCAG CTTGAGGGGC 240

TGCCTGGAGG TGGGGGTGGG GGTACACCTC CTCAGTGGGG AGAGACTTGC CAAAT 295

(2) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 370 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

TGGCCGATAT GAGGGGGGTG GGACTGGGCC CCGCGCTGCC CCCGCCGCCT CCCTATGTCA 60

TTCTCGAGGA GGGGGGGATC CGCGCATACT TCACGCTCGG TGCTGAGTGT CCCGGCTGGG 120

ATTCTACCAT CGAGTCGGGG TATGGGGAGG CGCCCCCGCC ACGGAGAGCC TGGAAGCACT 180

CCCCACTCCT GAGGCCTCGG GGGGGAGCCT GGAAATCGAT TTTCAGGTTG TACAGTCGAG 240

CAGTTTTGGT GGAAGAGGGG GGCCCTAGAA ACCCTGTAGC GCAATGGGGT TGGGCGCCCC 300

AAAGGTTAAG TTTGAACCCG AAGAGCAAAG GAAGAGGCGA TCATCATAAG TGGAGGATTA 360

GGATTAGGAT 370

(2) INFORMATION FOR SEQ ID NO:213:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ATCTGTGGAA TAATCTGCGG GCTAACACGG ATAACTCAGT ATAAGAACCA CCCAGTTGAT 60

GTCTATTGTG GCTTTTTAAT AGGAGGAGGA ATTGCACTGT ACTTGGGCTT GTATGCTGTG 120

GGGAATTTCC TGCCCANTGA TGAGAGTATG TTTCAGCACA GAGACGCCCT CAGGTCTCTT 180

GACAGACCTC AATCAAGATC CCAACCGACT TTTTATCTGC TAAAAATGGG TAGCAGCAGT 240

GTATGGGAAT TTTCTCATAC AGAAGGGCAT CCCTCAAACC GGAAACCACA GAGATGCTAG 300

GT 302

(2) INFORMATION FOR SEQ ID NO:214:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

ATGGATGAGT GGGCACCCCG CACAGGGCTG CAGGGTGGAA AACGCTCGAC GGCCAGGTGG 60

TGACTTGGGG GCAGAGAGCG CAGTGTNGTA GGGGAGGAGA GGTGGTGTCC CTGCTGCCTG 120

GGAGCCAGCC TGCCTGTNCT GTGGGCAGAG CAAGGCACTT TCTGCTGCCG GTGCTTCCAG 180

GGCCTAAGCA GCCGCTGCAC ACTCACCAGC GCAAGGCTCC TCTGCAGGGA ACGAGGGCTG 240

CTACCCATTT CACAGATGAG GGCAAGCAAG GACTTGCCCA GGGTTGCCCA NAGCAAGTGC 300

GTAACAGGCC CTGAGAAGAG NGCCAGTGAG CTCATCCTGA GTTAATTATG GGCT 354

(2) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

TGGTTCAAAG TCTAGGCCCT CTTNAGAGCT GGCTGATTCA GCTTGCCAAC AGTGACATCA 60

GGGTGAGGCT TCCTCTGTCC ACAGCATTAG CTGCGAATAT CCTCATGGTC ACAAGATGGC 120

TGCCAGTGGC CGTCAGGGTG TGTGCTTCCT TGTTCACATC CAGTGGAAGA GTGACAGCCT 180

GCTCCCCTTA GCTCTCTGAC ACCANTGTGA AGGTGCCANG AACTTACTAG CAGGNCTTTC 240

CTCATGACCC ATTCAACAGG 260

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

CTTGGACAAG ATCTGGGATA ATTCTCTGGA TTACCTGGCA GAGACTTTTK TTCTCTTGCC 60

TTACTGTCTC CCAAATAAAC AGTCTCTCAC TGTGTTGTGA GCCACCTGAA GCTGTGATAT 120

TTCCAACGAC TGTAGGAGGA AAAAATTAAG GGGAGAGAGG AAAACAAAAC CAACCAACCC 180

CTAANATCAT TTNTTTATTG TACATAACGA CCTCATTCTC CTGTATATGC GG 232

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

CTGCAACCAT CCATACCTTT TNCCCGTGGC TGCTATGGAG TCCCCCAAAC TCCCCAGTGG 60

GGCTTATGAG GGTGGGGCAC TTATTANGTN GTCTGGGAAG CTCATGCTGC TCCAGAAGAT 120

GCTGCGAAGC TGAAAGGAGC AAGGACACCG AGTGCTCAAT NTTCTCGCAG ATGACCAANA 180

TGTTAGCCTT GCTTGAGGGC TTTCTTAGNC TATGAGGCT 219

(2) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

GATAGGTAGC AGAGACCAAG GCGCAGGGTG CTTCAGATGA GCAAGAGAAC CCAGTCGAAC 60

CAGATACCCC AGGTGGGCCG GAGGGACCCC AGACCTTCAG AGGGCTGCCC TGGTGTTCTC 120

CACAGTGCAG TCCCTCTGTA TTCCCAGAGT GGGATCGGGG CTTTCAGCCC ACCCTGATGC 180

CTGCCCTCCA GGATGGCTGG TTTAGTCTGG GTCCATGTCC CAGACCCCTC TATTCTGCTC 240

CAGGACAGCA GGACTTCAGG TCTTTCCTGG GGGTGGATAT AGGAGAAAAT TTCTGCCTGG 300

CACACACCTG GGCTCCAACC ACTTGCCAAG TGATTCACTC TTAGGCCCAG GGGGAACACA 360

ATGACTATCA TTACTGATGC AGACCTGGCT 390

(2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

TTTTTGTTTT GTTTTAATAT TTTTGATATT CTCTTTGCAT TGAAATGGTA TAAATGAATC 60

CATTTAAAAA GTGGTTAAGG ATTTGTTTAG CTGGTGTGAT AATAATTTTT AAAGTTGCAC 120

ATTGCCCAAG GCTTTTTTTG TGTGTTTTTA TTGTTGTTTG TACATTTGAA AAATATTCTT 180

TGAATAACCT TGCAGTACTA TATTTCAATT TCTTTATAAA TTTAAGTGCA TTTTAACTCA 240

TAATTGTACA CTATAATATA AGCCTAAGTT TTTATTCATA AGTTTTATTG ANGTTCTGAT 300

CGGTCCCCTT CAGAAATCTT TTTATATTAT CCTTCAAGTT ACTTTCTTAT TTATATTGTA 360

TGTGCATTTT ATCCATTAAT GT 382

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

GACTTTGGTT TATTTAAAAA ACAAGCCAAA AAAAAAAAAA AAAAACCCCA ACTTTATATA 60

CAAAGTCAAA CTGAAACCAC GGWTTATGGA AAGAGGCAAG AWTTATGGGT AACAGGGGAG 120

AAGGCTGGGC CAGAGCCAAT ACCACATTCT GAACACAGGA GCCACGGGAA AGAGGTGCTG 180

GTTTCTTCTG GCAAGACCGG GGTGACTGGA ACGCAGTGGT CCTACTGGCA AACCCAGCCC 240

AACACTGAGC TCTTTCTAGC ATGGACTCCA TTCCCGTGAT TGGCCAAGGG AGACCCTTCC 300

CCCAGGAGGC CTGT 314

(2) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

TTCCTTCTCT GCGGCGGCAC GTCGCNAGCA GCCTGCTTCG CCCCGTCGTC AACTTTGAGC 60

TGGAGGAGAA GCAACTTTGG CAGTGGCCGC GGGGTGGGAA TCCCGCTTCT CCTCGGCAGC 120

AGTAGGCTCG CAAGTCGCTG GGGTTAGGTG GGGCAAGAGT TTCGCCGGGG CATCAGCGCT 180

TGCTTCGGAC TGTTTGCAAC GTGTTTCCAG CGAGCTGGGA GCGGGGGTTG TGACTGCGAG 240

TCGTCTGGGG GAGGGGGACT TGTTTTTCTT TTCCTCTAGA GACCTCGGCT TTCAACTGGA 300

TCAAACGTTG TCGAAAGGAT GTAAATAGGC AAGAGCAAAC TG 342

(2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

GTGATGGCTG CCTTGAGGGG GACCATCATG TCGGAGACGC ATTGGTGCAG GTCTCACCCC 60

ATACCCCATG CCCAGCCTCC TGCAGACTCA GGTCATCCAG GTGGTGGATG GCTCTTTGCA 120

TACCTGGTGC CTTCTCCTCT CGGGCTTGGC AGGCTTCTCT GGGGGCTTCT CAGATGACTC 180

TTTTGGCTTC TTCTCTGTCT TGGCTAACTC CTTGGCCAGC TCTCAACCTC CCTCCTTGGC 240

TCCCTCTTCT ACCACCTCCT CCCCTTTGGC CAACTTCCTC ACGCCCCTCT TGGTAGTGGC 300

TTTGAGGCTC TCCTTCCTAT CAGCCCGCTC TTTGATTTTC CTGGGCTTGA GGTTCCTAAG 360

GCACAGCCCC AAGAAG 376

(2) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 445 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

GTTGATAGAC ATTGGCATTG GGGTTGCTTC CACCTTTTGG CTGTCATGAA TAATATTGCT 60

ATGAACACTA ATGTACAATT CTTTGCCTGA ACGTAAATGT TTTCATTTCT CTTGGGTATT 120

TATCTAGAAA TGAAATTGCT GTATGTTAAC CCTTTGTTTA ACCTCTTGAG GAACTGGCAG 180

ACTTTTCCAA AGCAGCTGCA CCATTTTAAA TTCTAACCAG CAGTGTTTGA GGGTTCCAAT 240

TTCTCTATAT CCTTGGTAAC ACTTGTTATC TGCCCTTTTG GTTAGAGACA TCCTAGTGAG 300

TGTGAAGTGG CATCTCACTG TGGTTTTGAT GTGCATTTCC CTGATAGCTA ATTGTGTGGA 360

TCCCTTTTGC TTTTAGTGGA ATGAAATATC TGGTAGTCTC GTATGCCAAA CTAAAGCTAA 420

AATTAAAATG ACTCTGCATG ATGGA 445

(2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

TGCTCTCGGG ACAGTTTCCC GGGCAGCTCC TGGCCAGCTT CCAGCCCAGA GTCCTCAAGT 60

CCAGGGCACC TTGGGCCCAG CGCAGGCAGA ATCCGAGGTG GTCCTGGCTC TACCCTGGGC 120

CTCCTACTCC CCAGCACCCC TGGAGGAGGC AGGGGCTCCC CGCCGCCGAG GCTGCCTGCC 180

CTAGGCCCAC CTCTGCATGC TGCTCATGGG GCCACCCTGC CTCCTGGGCC CTCACTCTGC 240

CTAGGGGAGC TGGGCCAGGC ACTAGCCTTT GCCCAGGGAG GTGGGCCTCA GGCTGCCCAG 300

GTGCCTGCAC CCCAGCCGGG CTTCTCTGGG GCCTCCCCGT CGTCAAGCCT ATATCCTGTC 360

TGTCCCCACC CCAGCTGTCC CTTGCCAGGG GACTGGCATA AAA 403

(2) INFORMATION FOR SEQ ID NO:226:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 440 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

GTGCCTTAAG GAGAGAGATT GTGTTCTTCC TCTCTCAGGG GTGATAACTC AGGAAGCCTC 60

TGGGTTGGGA AGACCATCAG TTCTTTTGTC TTAGGTTTCT TTTCCTGTCC CTCTTCCATC 120

CCCAAGATGT GACCCCATAA AAATTTTTCC TGAGTTGGCC AGGCATGGTG GCTCACGCCT 180

GTAATCCCAA CACTTTGGGA GGCTGAGGCG GGCGGATCAC GAGGTCAGGA GTTCGAGACC 240

AGCCTGACCA ACATGGTGAA AACCCCATCT CTACTAAGGA TACAAAAATT AGCCGGGTGT 300

GGTGGCACAC ACCAGTAAGT CCCAGCTGCT CAGGAGGCTG AGGCAGGAGA TTTGCTTGAA 360

CCTGGGAGGC AGAGGTTGCA AGTTAGGCCG GGATTGCGCC GTTTGTACTC CAGCCTGGGC 420

AAGCAGAGCA AGACCATCTA 440

(2) INFORMATION FOR SEQ ID NO:227:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

GACGAAGAAG TTCCGGTTCG AGGAGCCCGT GGTTCTGCCT GACCTGGACG ACCAGACAGN 60

CCACCGGCAG TGGACTCAGC AGCACCTGGA TGCCGCTGAC CTGCGCATGT YTGCCATGGC 120

CCCCACACCG CCCCAGGGTG AGGTTGACGC CGACTGCATG GACGTCAATG TCCGCGGGCC 180

TGATGGCTTC ACCCCGCTCA TGATCGCCTC CTGCAGCGGG GGCGGCCTGG AGACGGGCAA 240

CAGCGAGGAA CAGGAGGACG CGCCGGCCGT CATCTCCGAC TTCATCTACC AGGGCGCCAC 300

TTGCCACAAC CAGACAGACC GCACGGGCGA GACGGCTTTG CACCTGGCCG CGGTTACTTA 360

CGCTCTGATG CCGCAAGGGC TCTTGAGGCC AGCGAAGATG CCAACATCAG GCAACATGGG 420

CCGAAC 426

(2) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

CAGGACCAGG AGAAGATCCT GGAAGATGCA GTGGATGAGT GGACGGGCTT TAACAACAAG 60

GTTAAAAAGG CCACTGAGAT TGTTTTAGAA AACCAACAGG AAAACACTGA CAAGCTAGAT 120

AAATACAGAT TGGACATTTT AGGCTAAATT CACTGTATTT CGTACTTGCT TGTAGGAAAC 180

CGAGTAAAGT CGAAAAGCTG TCCTGATGAT ATGGCATGCA CACCACACTG TAAAAAGTNC 240

TCCACACTAT TTAACAGGAC TGTGGCAAAA TAGCTTTA 278

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 425 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

TTTTTGTTCC CAAGCCTTTG TGACTGACTT TAAATCCTCT CACCTGCAGA ACAGAGATGG 60

CTTCAAAGTG GGGAGTGAGG GAGTGAGCGA GGACCCTGGG CTGAGACCTG TTTTTCTTCC 120

ATTTCTGCTG TGGCTTCCCA CAGCTCCCTG GTTCCACACC AGGCCCTGCT CTGCCGCAGA 180

AAATGGATTC CCAGGCCACA GAGCTGTCAG GCCTTTGACT TTGCAGAGAC CAAGCACCCC 240

AGAGGCTGTG CGACASGGCT AGTCCCTGGT GGGCCGGTCT GGGGCATGGG GGGCAGGGAG 300

ACTKGGAGAT GGGGAGGGCG TTGAGAATCC GGGGGGTCCT GGATACTTGA CAAATTGGCT 360

CAGGTCTTAG CTYTGGYTGC CCCACTGATT GTGTTGCTTG GCAAGGTGCA AGTYTTCGGC 420

TGTTC 425

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

TTGGAGGATG TGCTGCCCCT CCTGCAGCAG GCCGACGAGC TGCACAGGGG TGATGAGCAA 60

GGCAAGCGGG AGGGCTTCCA GCTGCTGCTC AACAACAAGC TGGTGTATGG AAGCCGGCAG 120

GACTTTCTCT GGCGCCTGGC CCGAGCCTAC AGTGACATGT GTGAGCTCAC TGAGGAGGTG 180

AGCCAGAAGA AGTCATATGC CCTAGATGGA AAAGAAGAAG CAGAGGCTGC TCTGGAGAAG 240

GGGGATGAGA GTTCTGACTG TCACCTGTGG TATGCGGTGC TTTGTGGTCA GCTGGCTGAG 300

CATGAGAGCA TCCAGAGGCG CATCCAGAGT KGCTTTAGCT TCAAAGGAGC ATKTTGACAA 360

AGCCATTKCT CTTCAGCCAG GA 382

(2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

GAGGCTGGAG AATCGYTTGA ACCCAGGAGG CGGAGGTTGC AGTGAGCCGA GATGGCGCCA 60

TTGCACTCCA GCCTGGGCCA GAGCAAGGTT CCTTCTCAAA AAACTTGGAA ATCTGTTGGG 120

AAGTAGGGGG AGGGCAAGGT TAAAACCTAT GCAGGTGTGT CAATTAGACT TGTTCCAACT 180

TGAGAACCTG AATTTTGCAT GTAATTGAAA TGTTCCAGAA CAAGTCTGGC AGTTTCATAA 240

GGGAGTTTTT AGATGCCAAT ACATTGCAGA TAACCATATT GGTTACATTA GGGGAATGAG 300

CATGGGATAG GTGCCTCCCA GTTGGTAGGA TAGCATGAGG AGGTTTCAAA AGTAACCSCT 360

TTAAGGGTTA TGTCCAGTAT TTGCTAAGTA ACCAAGGT 398

(2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 272 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

GGGGCTGCAG ACTGAGTTAT TTTATTTCGC TATTTCCAGT TTGAAGCTAC TATCATGGGC 60

GTTTAGAGTT ATACAAATGA CACTTACAAA AAATAAAAGA CCAAGACACC CAGAGTGAGA 120

TGCATGTTGG GGACGGGGGA GGCTGGCAGC AGGGGGGCCC CGGCGGYTCA CCCCAGGGCT 180

CCGGGAGGGG CGACGCCTGG CTTCATCCAC CCGGGAGGCC CAGGGAGCAC CAATCACAGC 240

AGGGGCTCTG GCCCAGGTGT CGGCAGCCCA GG 272

(2) INFORMATION FOR SEQ ID NO:233:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

ATTTTACAGT TTTATTTTTA AATCATTTAC ACATATTCAT ACAAAGAAAA ATAAATTTCA 60

GGATGGAATC CTGGGGACCA TGGTAGTTTA AAAAAAAAAA TCTCTCTGAT CATTAGCTAC 120

TAAAGACANG GCAAGAGGGT TAGCAGTCAT TTCTGGGGGT TAGTGTATCT CCCCATGCAG 180

GGGACAACTG NGAAGAATCC AAGCTGCTCC CTCATCTTCC TTCGATCTAG ATGGGGGAAG 240

GGGATTTTCC AATGCTCTCC CCTAGAAACA TTTCAAGAAG TACAGCAAAG GCTTATGGTA 300

ACACTGGAAC CTATTTGCTA GAAATCTGGC AAGATTGCAC TTTCTGAACC CAATTTTCCT 360

ATAA 364

(2) INFORMATION FOR SEQ ID NO:234:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 217 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

GGCCAGGAGC CAGAGGGCCC CGGGGCCACC CCTGCCGGGG AACGTGATGA CCAGAGTCCA 60

GACAGTGTCC CAGAGAGGCC GCGGCCCGCA GACCGGAGGC TCTGTCTGCC CTNCGTGGAC 120

GCCTCGCCAC TCCCAGGGAG GACGGCCTGC CCGTCGCTGC AGGAGGCCAC GCGGCTCATC 180

CAGGAGGAAT TTGCCTTCGA TGGCTACCTG GACAATG 217

(2) INFORMATION FOR SEQ ID NO:235:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

AACTTTAAAG TTAGGATTTT AAAATATTTG TAACTGGCTA AATTTTAAAG TCGTGACAAA 60

TAATTACTTA GGTTCAGAAA TATACACACA CTTACTCTTT AGCCAGTTTC TTTCAAGGTN 120

TTACTGTCCC ATCAGATATC TAGCCATTTK CCTTTGCAAA TTACATACCT TCTTAAGAGT 180

GTATTTTTAA GATTATTACT TATGCTTTAT GATGATATAG T 221

(2) INFORMATION FOR SEQ ID NO:236:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

ATAAATGGGT TTCTCACTCC TTAGGGACAC GATTGGAAAC AATACATCCC ATGAACACAG 60

GTGAATGTCC CTGGTTATCC CTGAGCTGGG CAGTTTCACA CAATCANTTT TNCTCTGAGG 120

CCAAAGTCTG TGGTTTGATC ATCTTAGCAG CTTCCAGAAC AGAAAGTAGG TTTACTTTGT 180

CTCCAAANTC TNATTCTCGG TGCTCAAAGA AGAATGACCT G 221

(2) INFORMATION FOR SEQ ID NO:237:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

GACATCTTTC TAAGATTCTC TGTGGGAAAA TGACTGTCAA TANAATGCGG GTTTCTGGGC 60

CATTCGTCTT ACTTTCATTT TTTGATTACA AATTTCTCTT GACGCACACA ATTATGTCTG 120

CTAATCCTCT TCTTCCTAGA GAGAGAAACT GTGCTCCTTC AGTGTTGCTG CCATAAAGGG 180

GTTTTGGGAA TCGATTGTAA AAGTCCCAGG TTCTAAATTA ACTAAATGTG TACAGAAATG 240

AACGTGTAAG T 251

(2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

GTTCGTGGCT GTCACAATAA TGCTGTGATA ATGCTGTGGT TTCCCAGCAG GGAGGTGGGA 60

GCGGGGAGGG GGCTGCAGCC TGATGAGAGC CAGCTGAAGG AAGAGCTGCC TCTCCCTTGC 120

TAAGCCGCTT CCCAAGGTCT GCCCCACCGC CCAAACCAAA GACCACTCCG AACAAAGTGA 180

GGATGTGGAT GCTCTTGCTG GGTCCGCTGT TCCGCAGAGG GAAAGAAAGG GTAGCTGCAC 240

TGACCCCACT GTCCCCATAT ACAAGGGTTK GGGGGCAAGA GCATGTGGCT ACTCCCAGCA 300

AGGGRAAAAT GGGAGGAGCA GTAGAAA 327

(2) INFORMATION FOR SEQ ID NO:239

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

ATTATTAGTT TATGGTGCTT TAAACCTATC AAAATAGTTG TAAGTAAATG GATTTCTTGT 60

NGTCCGAATA AGAATTCTCT GAGCTAGGAT AGATGTCTTT CTGCCCATTT TACAGGTGAT 120

GACACTGACA TAGGGACTGA GTGGGTAGCT TAAGTNCCAT GGTTACCAGG AGCAGGACCN 180

ACGTTTCCTG NCTCCCAGTC TCATCCTGTT TTCCACTCAC CAGGTTGGTT GCTCCCTTGG 240

AAAGCAGTCC CTGAGAGTTG ACTTAGAAGT TGARRGNGAA GAJGT 285

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

TTTTGCCATG TTGGACAGGC TGATCTCAAA CTCCTGGCCT CAAATRATCT GCCCAGCTTG 60

GMCTCCCAAA GYGCTGGGAT TACAGATRTG AGCCACTGCA CCCAGCCTGA CATGCCATAG 120

TTTCAGCATT TTCTTGGGCA ATGATCCAAG CTGAAGGCTG GTCTGAGGGA TCTSAAGAAG 180

CGTATGAGTT GGAAGAGAGG GACAGAAAGG AAGAAGACAT GTGAAGAGAG AAAAGGAAGG 240

AAGCTAGCAG AGGAATGCCC TCCAATAGAG ACTGCTGCCT GAAGCTCAGC CCCTCTGAAG 300

ATAGGTAGGC CAGGCTGGCT TAGCTGAGGC AGTGGGTTAG ACCAGCCCT 349

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

GTGCAGCGGT CTGCCTTCAT CTTTTAATGG CCGGTGCGGT ACAGTTAGTG GACAGACGGG 60

GGATGGGACA CAGCAGGGGT GAAACAGGGC AGTCACAGCC GGGGCCGGGG ATCTGGAAGC 120

GGGGGCGGTC CTCCCCCTGG AAACACCGTN TCTGGAAGGA CACCCTTAGG ATCCCCTGAC 180

CTCARGGTGC CACCCACACG GGCCTGGTGT TCTGGGAGGC CCGGCTKGAG TGA 233

(2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

ATATGTACTA CATTTGGTGG AATACGCATG TACAATTCTT CAAAAATAGT AAAGAGCAAA 60

ACAAACAAAA AATAGTAGAA GCACTGGAGA AATACACTAT GGCATAAACT AGTTACGGGT 120

GGGATGTCAC ATGGACCATA TCTACACTCT GTGGCAACCT TCTTACCTGA CTCCAAAGGA 180

TCAGATAATC AAACAGGAAA TTATGGTAGG AAATCAGAAA ATTGAAGTAT GCATTCATAT 240

CCTAAGCATT TTATTTTAGC TCAAAATATA AAAATATTCA TCAGTTAGCC AAGCTTTTGN 300

GATGAGAGAT CATAGCCTCC TCTTTGATAG GGGGTTTCTT GGGTTTCCTT GATTTCATGT 360

TTCAGAGTTT TT 372

(2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

CTCACACATT CATACCCAAG GAAGAGGCAA ACACACTCAA GTCCAGAGTT CCCAGTGGTG 60

CGGCCCAGAC CTACTGTCCC GGGGGTGTTA TGGCTGTCCC TCGGCTTCCC CAGAGCAGCC 120

AGGACAGCCT GCACCGNCTN CCAGACTCTC GCAGGAAGGG GAGCTCTGCC CTGGGGAGGA 180

AACTNACAGG CTGGGAGACA AGACTCCCAT CGCAGGGACA TGCACAGCAG CAGCCACAGC 240

CCCGCGGACG GGGCAT 256

(2) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 220 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

CAAATGGCAG TTCTCGAGAA TCGACGAGGA ACTTAAATCT GGACTCAGGG TTTCAGTGCG 60

GTCTCGGACT CCCACCACCC CGCCCCTCCG NCTGTCTCGC CGCCAGGNGT GACCTCCACG 120

CGAAGGAATC TTCTTCGGAT GGGTGCACCT TGCCAANAGG TGTGGCACCT GGNGGACTAG 180

GAGGCGCCTC CANACTAAGG GCGCTCANTG CGGCGTTCTT 220

(2) INFORMATION FOR SEQ ID NO:245:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TTCATGCTCA TGTAACCTTC TTAATAGTGC CTTGTCTGCT GGGTTTGTAG CTGTAAGACT 60

TGTGCAAACT GGCCCTATAA AAATATTGAT GCTGTCCATT AAAATGAATC TGTCTCTCTC 120

ACTCAGTCTC TCTCTCTGTC TGTCTCTCTT TCTTCTCTCT GCTGCCATGT GTGTGTCTCT 180

CTGTAGTCCT CTGATTTTGN CCTCTCTCTC TATTCTGCTA CTCTCTCTCC TCTCCTCCC 239

(2) INFORMATION FOR SEQ ID NO:246:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GGTTTAACCA GCCTTTAATA TCCTCTGATC TTCACCCTTC CTCTNACTNT TCTCCCCAGA 60

AGGGGGTGGG GCGGAGGGTC AGGAAAGCAG GCTCAGCTTC CAGGGTCAGG GAGTTGTGGG 120

CCCAGAGGGG CTGTCACAGT GGATGCACCC TGCCCCCTCC CTCGCCAGAC CCGAGGGTAG 180

GGCAGAGGCA CCTCCTCGNC AGCCTNTGGG CTGCACCCAC AGGGAATNGA GGGGAGGGGC 240

ACCATTACCA CTGGACCCAC CAAAGACCC 269

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 297 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

CTATTCAAAG TTTACTGACC TCCCCAGCCA GGCAGGCCAA CCCTTCCGAG CAGGGGAAAT 60

GTCCATCTAG CTGCCCTCTG CTGGGTTGCA GCCTATGCCA TGAGAGGGTA CTGGAAGCAG 120

GAGGGAGCCC TGGCTAGGGC AGGCCTTAAA CGCAAGGGAA GCTGAGCAGA GATCTGCACA 180

CTCAACCCCA TTTGATATTC TTCTCCTCCT CAGTCATGGC CAGCGTGTTG GTGACTAGAC 240

CGGTGCCAAT AGTCCGGTTG CCATCTCGCA GGGTGAAAAG ATGGCCTTTC TCTTAAG 297

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 281 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

ACAACAAGCA CACCAACTAT ACCATGGAGC ACATCCGCGT GGGCTGGGAG CAGCTGCTCA 60

CCACCATTGC CCGCACCATC AACGAGGTGG AGAACCAGAT CCTCACCCGC GACGCCAAGG 120

GCATCAGCCA GGAGCAGATG CAGGAGTTCC GGGCGTCCTT CAACCACTTC GACAAGGATC 180

ATGGCGGGGC GCTGGGGCCC GAGGAGTTCA AGGCCTGCCT CATCAGCCTG GGCTACGACG 240

TGGAGANCGA CCGGCAGGGT GAGGNCGAAG TTCAACCGCA T 281

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

AGCGCATCCA CACCGGGGAG CGGCCCTACC CCTGCTCCTA CTGTGGCAGG AGCTTCCGCT 60

ACAAACAGAC ACTCAAGGNC CACCTCCGTT CAGGCCACAA TGGAGGCTGT GGGGGTGATA 120

GTGAGCCATC AGGTCAGCCA CCCAACCCAC CAGGTCCCCT CATAACTGGG CTTGAAACTT 180

CTGGCCTGGG TGTCAACACT GAAGGTCTAG AGACCAACCA GTGGTATTGG GGAAGGGAGT 240

CGAGGGGGAG TTTTGTAAAT CCAAATCTCT GTGGNTTCAT GCTTTGTATA TGCTCACAGC 300

AGGGCACAAT AATCCAAGAG AAGGTCTGTG AGCCCCNATC CAACACCCAC AGTAATTATA 360

ATCTTGGCAC ATCAATGGAA TTT 383

(2) INFORMATION FOR SEQ ID NO:250:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 397 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

GTATCCTACC TTACAACAAT AATATCATGG GAGAAAATAGA AATAGCCTAG TTTGCTTCCA 60

ATAGAAACTG CTTTTAACAT GGGCTGTATA TAAAAATATT AAAGAGAAAC AAAACTGTAC 120

ATTTCCTCAT TGCTCCGCTA CAGACAACCC ATGTCATAAC CTTGTTGCAA ATATTTTTCT 180

CCTATAGCAG TAAGTACAGC ATTAGAAGGT GATTAGAGAG TCTGTTGATG AAACACAAAT 240

GTATGTTTTT ATTGATTTTT ACTTTAGAAC ACTACAGAGT TCCTGGGACC GGGGTGAANG 300

GGATTTAGCT GGGGTGGTTT GTGTGGGGGT TAAATACCTT CCCACTTGCA AGTGACTTGC 360

CTGTNCCCGC TGCGGGAATC CTGTTNCTTG GGTGGGA 397

(2) INFORMATION FOR SEQ ID NO:251:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 276 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

GGCCATAAAA GAAAGAGCCT GTTACCTATC CATAAACCCC CAAAAGGATG AGACCCTAGA 60

GAGAGAGAAA GCTCAGTACT ACCTGCCTGA TGGCAGCACC ATTGACATTG GTCCTNCCCC 120

ATTCCCGGNC CCTGACTTGC TCTTCAGGNC NGATTTGATT GGAGAGGNGA GTNAAGGCAT 180

CCACCAGGTC GTGGTGTTTG CCATTCAGAA GTCANGACAT GGACCTGCCC CCCACCCTTT 240

TCTCTAACAT TGTCCTCTCA GGGAGGCNTC TACCCT 276

(2) INFORMATION FOR SEQ ID NO:252:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

CCTGAACAGT CTGTTTCATT TGACTGTTTG GGGGTCTCCC AGTTTAAGCA AGATATTTAA 60

GCCTTATTTC TCTTGGCATG CTTGGATTCC CCAGTAAAAA AAACTCCTGC CCTGGGCTGA 120

CAATCAAAGT TCTGGGAACT AATATGGATA AGCAAGCTGG AAATGGAGAA GGCTATTCAC 180

TGTGCCTGGG TCCTACTGTT TTCTGGNTGG GAACTGCTTT TCCATTAGGC CTGGTGTGCC 240

CTGGAAGGGA NGAGCCTCTT GCAGAGACTA CAATCTTGGA TGGGTCCTTT GCCAAGTTTG 300

AAGGTAGGAA CCCA 314

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

GAACACTCTG CTCCAGCCAA GGTGGTGAGG GCAGCTGTTC CTAAACAGCG CAAAGGCAGC 60

AAGCCACAGT CCCACAAGCC TCAGCCTACC CGTAAACTGC CACCCAAGAA GGACATGAAG 120

GAACAGGAGA AAGGAGAAGG GAGTGATAGT AAGGAGAGTC CAAAAACCAA ATCAGATGAA 180

TCAGGGGAGG AAAAGAATGG AGATGAGGAT TGCCAGCGAG GCGGGCAGTA GAAGAAAGGA 240

AACAANCACA AGTGGGTTCC ATTACAAATA GACATGAAGC CTGAAGTGCC CAG 293

(2) INFORMATION FOR SEQ ID NO:254:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 413 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

CTTTTTCTTA ATATATTAAT ATTTACCAAG GCAAGACAGT GATTTATGGA CATTTAAATT 60

AGTTTAGCTT TGTTCTGCTG TTCTAAAACA TTGTGTACTG TCTGATAGAC TTTTAAAAAA 120

CAGTGCTTTT CCAGGATGAT TTATGATATG CAGTATTGTT TATAGATGCC CATGGCTTAA 180

CCTTGAAAAG TCAATTAAGT GACACAATTA AGAGAGATAT GAATAGTGGT AGAAAAAGCA 240

TGTACTCTGG ATAAGTGGGG GTAAATCTAG TATTTGTTAT TCCTGTCAGT AATATTGTCA 300

NTAGTATTTT TTAGAAGGTT TAATTTTTTT ATGGGTTATA AATTCATGTC ACTCTTCTGC 360

AATGGGTACC ATCAGTGGGA ATGCNGGAAT TATCCATGCT TTGGGGGTTA AAA 413

(2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

GGGTCCAGGG GAGAATCAAT ATATCTAGTA TAGTTTATAT TTGTACCTTC TCTCCTTAAG 60

AGTTACAGTG AGTGACTCTA CTCCTCAAAT GGAGCACCTC TCTCCAGGAG AGTAAGAAGA 120

TCACATAAAT AGAAAGTGAG CTTTGGACTC TAACAGACAT AGGTTCATAT TCAACTCTGC 180

TACTTAATAT CCATATTGGT TTGAGTTATT TAACCTTGAC AATCCACACT GTAAAATGGG 240

TAAATAATAA ATACCCTCCT CTCAGAAGTG TTACAAAGTT TATATGAAAT AATGTGCTTA 300

AAAAGCTGGG TACATAGTAG GAGCTTAGTC ATTGTTTATT TTCTCCCTCA TACCCATACA 360

TGNTTCATTC CTACTG 376

(2) INFORMATION FOR SEQ ID NO:256:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 241 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

GTAGAGATGG GCTCACTATK TTGCCCAGGC TGGTCCTGAA CTCCTGAGGT AGGAGGATCC 60

CTTGAGCCTG GGAGACAGAG GTTGCAGTGA GCCGAGATCA CGCCACTGCA CTCCTGCCTG 120

GGTGACAGAG TGAGACTCTG TCTTAAACAA AACAAAACAA AAAAAGCCCA GGCCCAGGGC 180

CTCACACCTG GTAATCCCAG CACTTTGGGA GGCCAAGGTG GGTGGATCAC CTGAGGTCAG 240

G 241

(2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

GAAGGGTGTG CTTCGCAGGA TGACTGTTAA TGATTTGCCT GTGGGACGCT GCGTGGATGA 60

GGCTCTGCGG CTGGTCCGAT TAAGAAAACC AAGAGAGGCC GGGCACGGTG AGTGACGCCT 120

GTAATCCCAG CACTTTGCGA GGCCGAGGTG GGCGATCATG AGGTCAGGAG ATTGAGACCA 180

TCCTGGCTAA CACAGTGAAA CCCCGTCTCT ACTAAAAATA CAAAAAAATT AGCTGGGCAT 240

GGTGGCACGC GATTGTAGTC CCAGCTACTA GAGAGGCTAA GGCAGGTGAA TCGCTTGAAT 300

CCAGGAGGTG GGGGTTTCAA TGAGNCCGAG ATCGTACCAC TGCACTCCAG CCTGGGGCAA 360

CAGAGTANGA CTTCGTAACC CCCAACCAAC CCNCCAACCC CCCGCC 406

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 157 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

GAAAAGAAGG AAGGAAAGAG GGGAGGGAGG GAGGAAAGGA GAGAGGGAGG GAAAGAAGGA 60

GAAAATGCTG GAGCAAAGGA GGTTGGTTAC ATGATTTCTC TAATGGCAAT GAGCTGCTTT 120

CTGGATGAAA TACAGAATCA GAGCGAGACT CCGTCTC 157

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

AAGCAGATAT AAATGGGACC ACTGTGAATC AAAGGGGAAA AATTCCAGGA AAAAAAAATT 60

CCAATAGCTT CACAGTTTAA CTGAGGTTTT GGAAAAACTT AAGTGAATTC AGCTGATGTT 120

TGAAATATCT GTCTACATTT AATTAGATGT GTTGTATTTA CCAAGGAGGC ACAAATATGT 180

AGTTCTGTAG ATTTTAATAC TAACTTTTCC AGTAAGAAAA ATAATACCAG GTGATTTCAA 240

AAAGGGCAGT GATCTATAAA CACTCAAAAT GCATCTTTGA ACAGGGGAGC AGAAATAGCT 300

AATTTAATGA AAACAAACCT TAAGCACTTT ACTTGGCTTC TAATAAGGCA TCCCAAGAAA 360

A 361

(2) INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

CAATACATGT ATACAGTGTA CACTGATCAA ATAAGAGTAA TTAGCATATT TATCACCTCA 60

TTTCTTTTGT GGTGAGAACA TTTAAAATCC TTTCTTTTTG CTATTTTGAA ATATACAGTA 120

CATTGCTATT AAGTATAGTC ATCTGGCTGT GCAATAAAAC ACCAGNACTT ACCCCTCCTG 180

TCTGTGACTT TGTACCCTGT TCACCACCCC TCCAATCCTC TAGTAACTAC CATTCTACTC 240

TCTACTTCTA TGAGCCTGAC TTTTTAAAAT TCCACATGTA AGTGAGATTA CATGGTATTA 300

TTCTCTCNGT GGCTGGCTTA TTTCACTTTA ACATAATGTC CTCTAAATT 349

(2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

GGAAGATGAG GATCTAGGTG TGAGCGTGCA GAGCCCTGAG GCTGGGCAGG CAGGGAGCTC 60

TGCCTGCACA ATGATGTAGC CATGTGTGGC CACACCAGCA CTGGGCAGCA CCTCTGGGGA 120

GGGGGGCAGG GCAAGGACAA CTGGAGAGAC AAAGCCAGAT GGGGCCACGT CCTTAGAAGT 180

GTGTGTGCAC GCACATGTGT GTGTGTGTGT GTGTAATACG CAGGGCAGAA ACACACCATG 240

TAGGTCAGGC AGGACAGAAA CACATCATGT AGGCCAGGCG TGGTGGCTCA GGCCTGTAAT 300

GCCAGCACTT AGGNAGGCCA AAGTGGGCGG ATCACCTGAG GTCAGGAGTT CGAGACCAGC 360

CTGGCCAACA TTGCAAAACC TCATCTCTAC TAAAATTCTA AAATTAGCCA GGCGT 415

(2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

GGCATGGGGT CTGGCTTTAA TGTGTAACTG ACGTGGGTCA CTGAAACTGT TCAGGCTGAT 60

CTTGAACTCC TAGGCTCAAG TGATCCTGCT GCCTTGGCCT CCCAAAGTGC TGGAATTACA 120

GGAATGAGTC ACACCACCCA GCCGGCTGTG TTTTGTTTTT TGTTTTTTAC CCCGACAGGT 180

NCTCAGTCAC TGGTTAGCTG GAGTGAAGTG GCGTAACACA GCTCACTGCA GCCTTGATCT 240

CCTGGGCTCA AGTGATCCTT CCATTTCTTC CTTCCAGAGT AACTGGTACT CCAGGCCCAC 300

GGCACCACAC ATGGCTAATT TTTAAATTTC GTAGAGACGA GGTCTTGCCA TGTTTGCTCA 360

GGCTGGAGCT CTTGTATTCT TT 382

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

TGTATCAACT CAGAATTTCC AGAGAGCTCT TCCTGGCTGA AAAGATGTCC AAGGATCATC 60

TCCGGAATGG AAGAGGTGAG GCCTGTTAGC TTGTGGGCTG CCCAATCCAT CCAACCCTTG 120

GCATTGGGAT CAATGTTGAT GAGGACAAGA CCTTCAACAG TGTCCGGGTG GTTAAGAGCA 180

TATCTCGCCA GGATGTAGGC TCCAGCTCCA ACACCAACTC CAATTATTGT AGAGAAATTT 240

AGGTACTGCA GGACGCAAGG GATCATGTCT GCAAGCTGGT CCAGAGATGG GTACTGATAT 300

CCCAAAGGGA ACACAGGGGC TCCCTCTTCC ATTCCAGGGG CATCCACATG GACCCGCACA 360

AAGTTCTGAA TGATTTCCTG CATGTCCTCG AACTKGAACA GTGGCTGGAG GAAAGATTTA 420

TAGTTGAGTC CACATCGGGT AGGTAAG 447

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

TTTTCGCTGT CAACAGACAG TTTATTCTAT ATACAAACAC AATTTTGTAC ACTGCAATTA 60

AATAGAATGG AATGAGCGCT CCTCCGCATT CCTCCCCGAG TGACTGGTTT GGCCGCCGGC 120

CACTCCATCC CCGAGTGGGA CTGGACCACG GCCCTGGNTG CTGCCACTGA TGTTGGNGCC 180

TGCACCCCAC GTCCCTATGC CCGAGGCGCA ANTCTGCTCT CCCGGGGACC CCAAGNCTGG 240

NGCACACGCG GGGAGGGCGG GGCCATGGAG AAGGCACTGC AGGGAGCACC AGGCAGAGCC 300

GTGTTGAGGC CGGCCGG 317

(2) INFORMATION FOR SEQ ID NO:265:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

GCAGAGCAGG TGGAAGTGAT CAGGAACCAT AGTTGACAGT TCCAATCAGT AGCTTAAGAA 60

AAAACCGTGT TTGTCTCTTC TGGAATGGTT AGAAGTGAGG GAGTTTGCCC CGTTCTGTTT 120

GTAGAGTCTC ATAGTTGGAC TTTCTAGCAT ATATGTGTCC ATTTCCTTAT GCTGTAAAAG 180

CAAGTCCTGC AACCAAACTC CCATCAGCCC AATCCCTGAT CCCTGATCCC TTCCACCTGC 240

TCTGCTGATG ACCCCCCCAG CTTCACTTCT 270

(2) INFORMATION FOR SEQ ID NO:266:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 297 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

ATGAGGCGAG GCCTGCGAAG TGGCTGGCAT GCAGCAGGTG CTAATGAGTG TTGCAAAGGT 60

GATGTCACGC AGGCAGCTTC CCGTGGCCAG AGAAACATTG CAGAGAAGGG ATAAGTAGGG 120

CTTAGTGACT TTGACGGGTC AATGGAAGAA TGACCCAAAG AAGGCTTCAA GGCCAGGCCT 180

GCAGTTCTCC ACCACAAAGG CCCTCACTGA TAGCACCCAC TCCCCCACAC TCAGCTTTNG 240

GGCCTAGGTC TGGGTCACCC AGCTAGAAGC CACAGGACCC TGAGGCGTCC GAGGGGT 297

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 387 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

CTTGTTTTCA TCATGAGCTC GATCAGATGT CTCTCGATCT TCAGACTGGT GGTGTCCTAT 60

AATGTCCTGT GCACGCATTC TTGAGCTTTC CAGGATTTCT GTCTGTTCTC TCTGTTTATC 120

TACAGAAGAA ACTTTCTCCT TGAGTTCCTG TTCTTCGTAG CGCCTTGAAC TCTCTTTCCT 180

TTCTGGTTTA CGATCCTCCT CTTTCCATCT ACCCTGTCTG TCTTCTGTGA GGTGCGAGGG 240

ACTAAGAGAA CGAGATTCTT GAGGTCGTAC AACTTGGCTC AAGAGTCTGT GTTTTTTCAT 300

TTNTNATCAT CTCCACTGTT GTAGGCATCA CTGTCCGGAG AATGTTCACG CCGGCGCTTT 360

CGGGGGACTG TCTAGGGCTG GGACTCC 387

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

CCTCAGGTT ACTTCTTTCC AGAGAACATC [illegible] 60

[illegible] 120

[illegible] 180

AGGACGGCGA CAAGCCCCGG GTGCTCTACA GCCTGGAGTT CACCTTCGAC GCCGATGCCC 240

GCGTGGCCAT CACCATCTAC TTCCAGGCAT CGGAGGAGTT CCTGAACGGC AGGGCAGTAT 300

ACAGCCCCAA GAGCCCCT 318

(2) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 422 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

ACATGTCTAT TCAGGTCTTT TGCCCATTTT GAAATAGCAT TGCTTGTTCT TTTGCTGGAT 60

ATTAACCCCT TGTCAGGTGC ACAGTTTGCA AGTTACCTTT TCTCATCCTA TAGGTTATCT 120

CCTCACTCTT GATTGTTTCT GTTGCTGTGC AGTAGCTTTT AAGTTTGGTG TAATACCATT 180

GTGTTTTCTC TGCTGCCCTT TTAAGTTTCA CTGGGTCAAA AGTTTAAAAT TTGTGAATTC 240

CTATATTTTT AGGGCAATTC TCCTGCCACT GTTGGAATTA TGCCTCAATC TATGCAGTAG 300

AATATTAGTG TGAAATGCTT CTGTACCAAT GGAGATGATG CTGGATGGTC TCTATCATAA 360

ACCCATACCT CATCAACACA AACTGCAATT ACACAAGGGC TCTATATCAT GGATCTCCAT 420

TT 422

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

GAAGAAGAGC CCAGACCTAG GGGAGTATGA TCCACTTACC CAGGCTGACA GTGATGAGAG 60

CGAAGACGAT CTGGTGCTTA ACCTGCAGAA GAATGGAGGG GTCAAAAATG GGAAGAGTCC 120

TTTGGGAGAA GCGCCAGAAC CCGACTCAGA TGCTGAGGTT GCAGAGGCTG CAAAGCACAT 180

CTTTCAGAAG TCACCACGGA GGGCTACCCC TCAGAACCCC TTNGGGGCCT GGAACAGAAG 240

GCGGCCTCCT CCCTGGTGTC ATATGTGCGC ACGTCTGTCT TCCTGCTTGA CTTTGGGGAT 300

CTCGATGATC CTGGTGCTCC TGTGTGCTTT CCTGATCCCC TGTCCTCCCA GAGATCTTGA 360

CAGAACTGGA GCCGCA 376

(2) INFORMATION FOR SEQ ID NO:271:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

TGTTCACGTT CCCTTTCTTT GTCTTTCTTT TTCCTATCTT TATCTATACT TCGACTCCTC 60

TCCTTTTTCC TCTCTTGTTC TTTAGCCTCA CCTTTATGCT TATGACTGTN CCCACTAAGA 120

TTTCCACGTT GATCATCAAT TTTACGNCTA TCTCGACTCC TACTGCGACT GGCACGATTG 180

GTTCGTCTAT CCCTTGAGCG ACTTCTACGA ATGCTTATGA AAAAGAATCA AGTTGGNCAC 240

CAAATGTTTC ATAGCAGTAG GAAATTTCTT TTAGAGACTT CTGATGGGAA ATTTGAAGTG 300

TATGTTGCTA TCAGATCAAG TGCAGGAGAG GTATAAGGCT ACTGGA 346

(2) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

GTTGTTGTTG TTGAGTCGGA GTCTCGCACT GTTGCCTGGG CTGGAGTGCA ATGGTGCAAT 60

CTCGGCTCAC TGTAACGTCC GCCTCCCAGG TTCAAGCGAT TCTCTTGCTT CAGCCTCCTA 120

GTAGCTGGGA TTACAGGCAC CTGCCAGCAC ACCTGGCTAA TTTTTTATAT TTTNAGTACA 180

GACAGGGTTT CACTATGTTG GCCAGGCTGG NCTTGAACTC CTGACCTTGT GATCTGCCCA 240

CCTCAGCCTN CCAAAGTTTT TCAGAATTTT TTAAGGAAAC ACTTTTAACC CTTAAGGCTT 300

TCTTTCAAAC TCAGATCCCC TTACACAATT GATCAGACGT GGCAAAGTTT TGCTTCAAAG 360

TTTTTGGACT GGGTTTCCAC TTTAGGCTTA CTGA 394

(2) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 259 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

CAACCTGTAC CCAGCCTCCC AGAACGTTAC TTTTAGGAGC CCCAGCATGA TGTTCCAGCC 60

CCCTCTTTAC CAAACCRATC TCCACCTCTT TNTCCCCCCC ACCCCACCAC CCCACTCCTC 120

CCCCCATCAC CACTCCACCA ACCCCATCCA CATCCACAAC CCTTATTTTA ATCCACTTCC 180

CCATTTCACC CTCTCCCACT TCTCTCCAAA TCCTCTCTAT TTCTCCCTCT ATTACTATTT 240

TGCTGCAAAT AATCCCACG

(2) INFORMATION FOR SEQ ID NO:274:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

TCCCAGTTGT CCCGATTGTA ACTCAAAGGG TGGAATATCA AGGTCGTTTT TTTCATTCCA 60

TGTGCCCAGT TAATCTTGCT TTCTTTGTTT GGCTGGGATA GAGGGGTCAA GTTATTAATT 120

TCTTCACACC TACCCTCCTT TTTTTCCCTA TCACTGAAGC TTTTTAGTGC ATTAGTGGGG 180

AGGAGGGTGG GGAGACATAA CCACTGCTTC CATTTAATGG GGTGCACCTG TCCAATAGGC 240

GTAGTATCCG GACAGAGCAC GTTTGCAGAA GGGGGACTCT TCTTCCAGGT AGCTGAAAGG 300

GGGAAGACCT GACGTACTCT GGGTTAGGTT AGGACTTGCC CTCGTGGT 348

(2) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

GTTTGGTGAA TTTGGTCTGT GATAAAATTG GAGTTCAAGA AACAAACAGG AAACTACAAG 60

TGCCCCTTCG CCCCCAGGTC ACCCGAGTGG CAGGGCAGTG ACCGCTGCTC TCAGGCTGCC 120

CAGTGTGGAC CTGCCTGTCG GAATGCTCCT CCTCCACGTC CCCTCGCTCC TGTGTCCCAG 180

CCACATGCAC CTTCCCTCTA CCTCTGGGAT CCCTGCACCA GGTCTGCCCC TGTCTTCTCA 240

GGGCTGCTCC TNTTGGNCCA CAGGACCTCA GCTGGAATGT TGCCTCCTCC AAGAGGCCTT 300

CCTGACTATT CAGCTCACAG TGGCCACCCA GCCACAATCT GCCATGTGCT TTGGGGGATT 360

GTCTGTTAAC TGGCAACATA CTGGCAGCCC ATAACT 396

(2) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

GGTGTCGGGG AGGCTGCGCA AGGGGGCGAG CCCGGGCAGC CGGCGCAACC CCCGNCCCAG 60

CCGCACCCAC CGCCGCCCCA GCAGCAGCAC AAGGAAGAGA TGGCGGCCGA GGCTGGGGAA 120

GCCGTGGCGT CCCCCATGGA CGACGGGTTT NTGAGCCTGG ACTCGCCCTC CTATGTCCTG 180

TACAGGGACA GAGCAGAATG GGCTGATATA GATCCGGTGC CGCAGAATGA TGGCCCCAAT 240

CCCGTGGTCC AGATCATTTA TAGTGACAAA TTTTAGAGAT GTTTATGATT ACTTCCGAGC 300

TGGTCCTGCA GCGTTGATGA AAGAAGTGAA CGAGCTTTTA AGTTAACCCG GGATTGCTAT 360

TNAGTTAAAT GCAAGCCAAT T 381

(2) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

TTAATACGAC AGGGCTGGCG CCCGAGTAAT TCAAGCCCTT CGGAAGTGTC ACCGGCTGCC 60

AGGCCTCGGA TGCAATCCTG GAGGCGGGAG ATTCGGCCTN AAGACTGGCT CGAGCCGCCC 120

AGGGGCTCCA TGGGAGACTA ACGCGGAAGT YCCAGCCGTC CCAGTGCCGT GACGTCCCCC 180

CTTGGTGGGG CCTGCACCCG ACTACT 206

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

ACCTGTAATC CCNGCACTTT GGGAGGCTGA GGTGGGCAGA TCACGAGGTC AGGAGATAGA 60

GACCATCCTG GCTAACACGG TGAAACCCCA TCTCTACTAG AAAAATACAA AAAATTAGCC 120

GGGCATGGTG GCGGGCGCCT GTAGTCCCAG CTACTCGGGA GGCTGAGGCA GGAGAATGGC 180

GGGAACCCGG GAGGCGGANT TGCAGTGAGC TGAGATCGCG CCGTCTCTCC AGCCTGGGCA 240

ATAGAGTGGG ACTCCATCTC 260

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

CTCTCTGGGC TGAGGCTTGC TGAGCTTGGA GAGGAGCAAG GTAGTAGAAA TATTGCAGGG 60

TTCCCAAAAC CAGGTCAACC AAGATGCCAT GTGAGCCCTG AGCATGCCTG TCTTCCCAGG 120

GGTGTACCTC TTGGCTGGCA AAGCCAAGGC CAGTGGGNAC TTGTATAAAT CACATGGGTA 180

TGTTCTTGGT TCAGTGATCT TGGAGTGATG ATGGTAACTN ATGAACAGAG AACTTTYYAG 240

AACTTKGGTC CTGTCTTCCT CCCTGAACCT AGACAAGTTT CACCCCTCCT CCTGTACCCA 300

ACCCCATT 308

(2) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ATTTTAGCAG CTTTCTTGAA ATTTAAAATA TATGTGTAAG TATCTCATTT ATATGCATTT 60

CTAGTTTCTT TATACAACAG AATAACTTCT TTTACATCAA ATTTCTGAAT TTGACTAAAT 120

TTAGAAATAA TGGAATCTCA TCCATTAAAT ATAGTCATAG AAGGAAGGAA ATATGAAAAT 180

TAGGATTTCA GATGTTTGAA CATAAAAGAT AATTTTAAAC ATTGTCAGTA ATCTATTTCT 240

TTTTTTTTTC GAGACGGAGT TTTGCTCTGT CACCCAGGCT GGAGTGCAGT GGCGCGGTCT 300

TGGCTTACTG CACCCTCTGC CTCCCAGTTC AAGTGGATTC TCCTGCCTCG NCCTCCTGAG 360

TAGCTGGGGT TACAGGGGCA TGCCAACATG CCGGGGCTAA TT 402

(2) INFORMATION FOR SEQ ID NO:281:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

GAGAATCCGT CTTAAAAAGA AAAAAAGAAA ATTATAGAGG GAGATGAGGT GGGACAGAGT 60

CTGGCAGTTC ATCAGGGGGA CTGAGAAGGT GGCATTTGGA GGAGAGGAGG CAGTGAGCTG 120

TGCAGTGTCC AGGCAGCCAC CCTTCCCAGC GGCCACCATG ACGGTGTCCT CATTGCTTTA 180

ACCATTAGTA ATCATTCATT CATTCATTCA TTTATCCGAC GTCAGCTGGA GGNCCTGCCC 240

GNGGGGCATG CGCTTAGATT TNGGAGGCCT TCCGGGATGC TTGCGCTCCA ACGGGGGAAG 300

GCCGACTTGG GCT 313

(2) INFORMATION FOR SEQ ID NO:282:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

TGACCTCAGT TGATCCACCC ACCTTGGCCT CCCAAAGTGC TAGTATTATG GGCGTGAACC 60

ACCATGNCCA GCCGAAAAGC TTTTGAGGGG CTGACTTCAA ATCCATGTAG GGAAGTAAAA 120

TGGANGGAAA TTGGGGTGCA TTTTCTAAGG ACCTTTCTAA CANATGGCTA TAATNTAAGG 180

GGTTTAGGGT CCTTTTTTTT TTTTCAGGGA TACATTT 217

(2) INFORMATION FOR SEQ ID NO:283:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

TAGAGAGCGC TTTACTCCTG GTCCCATGGC GTAAAGATGT GGCTGGGCCT GACAAGGCTC 60

AGCCTCCAGT CTTAAGATGG GCACAGAAGG GCAAGAAGTA AGATGACGAG TCCCAGAATT 120

AGGACAAGCC ATGAGCCAAG GCCTGGTCTG AGCAAGGGCA GCCCCCTGTC CCAGACACAG 180

GCACCCCCAA TCTCACTTTG GACAGAGCCA ACGTGGGGGG ATCCTCCCGG GCCTGGGCCT 240

GTCAAGTCTG CCTGCAGGAC CCTGCCATTG TGCTCAAATC ACAACCATTT TTTGCTTCCA 300

ACATTTTAGG GTGCTTGTGC AGTGAGT 327

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

CTTTCGAAAT GTAAATTGTT ACAAACTTAC TTTAGAGCAA ATTTAGTCAT CCTTCAAAAA 60

TTTAAATGTA TACTTATTTC CTAAGAATTC GTTTGGCTCA CACAATTGTG AAAAGATAGA 120

TGTACACCAC TGTTCATTAC AACAATTATC CAACAAATCT ATTATGTGCC AGACATTATT 180

CGGAACTCTC GGAATACATA AGTGAACAAA GCAGATTCCT GATCTCAGGA CCTGGGGTCA 240

GGGCTCAGGA GAAGCCAAAA AACAGGCTNG AGAAATACTT TATGCAGTGT GGGGGGAGTG 300

CTACCAGCAC AGCAGGGGAT GGNGATGTGA AATCTTGTGT 340

(2) INFORMATION FOR SEQ ID NO:285:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

GACATTCACG GAGGTGGGTT CGACCTCCGG TTCCCCCACC ATGACAATGA GCTGGCACAG 60

TCGGAGGCCT ACTTTGAAAA CGACTGCTGG GTCAGGTACT TCCTGCACAC AGGCCACCTG 120

ACCATTGCAG GCTGCAAAAT GTCAAAGTCA CTAAAAAACT TCATCACCAT TAAAGATGCC 180

TTGAAAAAGC ACTCAGCACG GCAGTTGCGG CTGGCCTTCC TCATGCACTC GTGGAAGGAC 240

ACCCTGGACT ACTCCAGCAA CACCATGGAG TCAGCGCTTC AATATGAGAA GTTCTTGAAT 300

GAGTTTTTCT TTAAATGTGA AAGATATCCT TCGCG 335

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

GCACAATTAT TAAAAAGAGG CCACTTAAAT TCAACTCTCC ATGGATACAG TGTCTGTGGC 60

AATGTTTAAT TAGAGATTAA AATTGAGGAA TTGAATAATT GAGGTTGCTA ATGAATTTGA 120

AAACTCAGCA AAGCAAGGAG AGCTGAGCGT TTTTCCGACT TAGCTTTTCT TTCTCTAACC 180

CTTTTCTCAT TTCCTACTAT TATCACATNT CTGGCCTTGA CTGCTGAGTT TATTACTACC 240

CATAACCCTG GCCTAAGTGG AAACAAAAAA GCTGTAGCCT CTTTGCTGAG CTCCTGGAGA 300

CATTTGGTCT ATTGGATTTA TGACATGTTC AGAAGCTTGC AGTTGCAGGA GGCTGACAAT 360

GATGAAAATG AGATATGNTG GGCCACCACG CTTTTCTGT 399

(2) INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

TTCCAGTTGA ATTCACCAGT GGACAAAATG AGGAAAACAG GTGAACAAGC TTTTTCTGTA 60

TTTACATACA AAGTCAGATC AGTTATGGGA CAATAGTATT GAATAGATTT CAGCTTTATG 120

CTGGAGTAAC TGGCATGTGA GCAAACTGTG TTGGCGTGGG GGTGGAGGGG TGAGGTGGGC 180

GCTAAGCTTT TTTTAAGATT TTNCAGGTAC CCCTCACTAA AGGCACCGAA GCTTAAAGTA 240

GGACAACCAT GGAGCCTTCC TGTGGCAGGA GAGACAACAA AGCGCTATTA TCCT 294

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 391 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

TCTACAGATG AGGAAAGCAA GCCTCAAGCA AGGGGGGCCT GATCCTTTCC CTGTTCCCTG 60

TGTATTCCCT GTCTGTGGCA AAGCCCATTG CCTTGATTCT CTTCTCTTTA CTTTCATGTT 120

GAGAAGTAGT TTCTTTCTGC AGTTTATTTA ATTTACTGGC AAAATGACGT ATTTTTTTTT 180

CAGCAATGTT TCAGGTAGAT ATTTGCTTTA TGCATGTAAT GTCAATGAAG TACTCATAAG 240

TTTTCAAGAA ATGACTGATA TAAATCATGT GTTCCACTAC ATAGTCTAAA TATTTAGTAT 300

TTGGTCATCT ATTTTAATAT GTTCAAATTC TGTTAAACAA GNCATAGTCA CTATGTGAAG 360

ATAAAAAATA GNAAAGTTGC ATTATGACTT T 391

(2) INFORMATION FOR SEQ ID NO:289:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

GTTATATTCT ACTTTATTTG GTAAAACTCA GAAACTAACA ATTCACATCC TGCCACCTTC 60

TTCTTTCCGA AGAAGGCAGT TTGCAGAGAC AAAAGGGCTG TGGCGTGGGG ATCATCCACC 120

ATCTCCAGGT TTTACACCCA GGCTACCCAT GGCTTGGCAG TCAGGCCTCT AGGCTGATTG 180

CTCTCAGAGG CAATAGAA 198

(2) INFORMATION FOR SEQ ID NO:290:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 353 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

GGTTTTCATC TTGGGTTTAC AAAAGTGCTA GTATTTATTT ATTTTAAGTT TAATTTAAAT 60

ATCAGCTACC TTAGGTAGAA GTTTTCCTTT GTGTAATATA ATATAAAACC GACATTTCTT 120

GGGGGCATAA TAGTAAAGAT GTTAACATTT TTTGCTTGTT TTTGGATGGT GTATTTGTGC 180

TTGTTGTGAA AGTGATGTGT GCCAAGATGG GTCATGTAAC GCAGTTTTGA GTAGGGTATT 240

GATATTGTGT GTGGTTAATT TATTGAAGTG GGTTAAAGGT ATACATATTT GGTTTTAGNT 300

AAGTATGTAA GATATTGTAG ATATATTGGT GTAGTGATTT ATAATATGAT TGG 353

(2) INFORMATION FOR SEQ ID NO:291:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 163 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CCTGGTAGGC CTGCTACACA GTCTTGCAAC GNCCCTCGTG CTTGGGCTTC TGCGGTGAGG 60

CAGGGGAGTC TGCTTGTCTT AGATGTTGGT GGTGCAGTCC CAGGACCAAG CTTAAGGAGA 120

GGAGAGCATC TGCTCTGAGA CGGATGGAAG GAGAGAGGTT GAG 163

(2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 397 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

ACGGGAAGGT GAGTATGTNA GTATGTNTGC CAGACAATGG TGTTTCCATG TCAATGGAGG 60

TTTCTCAGAG AGAGGTGATC TGGCTGGAGA AAGCTTAATC TGGTGGCAAT GGACAGGTGA 120

CTTTAAGAAG TGGGGAACGA GGGAAGGAGG CCAGTTTGAA AATNATAACA AGGGTCCAGA 180

CTCAGTGATG CAGCAGTGAC CATGAGAACA GAGCAGCTGC AGGTAGAAGA TGGAGACAGA 240

ACTNGGGAGA TCTGGTGGAG GTAAGCCGCG TGGAAAGATG ATGTCAGGTT TATACCTAGA 300

GGACACATGA TCCATTCACA AAGCCAGGGG NAACCTAAAG AGAAAACACT TAGAATTTTN 360

GGAGAANAGG CTAGGGCTGG GCCTTAGACA TGGGCTG 397

(2) INFORMATION FOR SEQ ID NO:293:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

GAGGTAAAAT TTACATACAG TGAAATCCAA ATCTTAAGTG TACCACTAGA TAAATTTTGA 60

TAAATGCATT ATGCCTGGTC TTCACACACC CTTTTCAATA TATAGAAAAT NTCCAGATAA 120

TTTATTTTGT TGTTTTTTTC ACACACTAAG TTCTAGACTT TTCCAGGTCC GAGGGAACTA 180

TTAGGGGGGA AAGTACTTGT NATAGTAAAA AAGATTTTAG GTGTGTTTGT TTTTAAGGTG 240

CAGAAACACA TCGCAGATTT AAGGTCTGCA ATCTCTGCTT TTTGTTATTG TTCCAGTTTT 300

GATCTCAGTG ACATTACAAG CAAGCAGAAA CACTCAGACA TGAAATGGCC CAGTGCCTGT 360

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 321 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

TTTTTTTCAG GNTTCAACGG TTTTATTGGG AGGTTTTGTT TTCTGTGAAA TACACTAGAG 60

GGTGGGGAAG GGGACACATT CACTTTGCAA GATAAGGGTT TCCCACCACT AAAGGAAAGG 120

CATGGGGCAG GGCACACTGG GGTTTGGGTC CGTTTTCCCA CCTCCTTCTG CTTGGCTCAC 180

TTTTCTTTTC TCTCAGCAAG TACCACAGAA CACAAAGACA AGAAACAAAA CAGCAAATCA 240

ACCTCCAACG GGGCCATGCC AAGCCTTCCC CACTCCCCCA GGCTGGGCAA GGGCTGGGAG 300

GGGGCTGGGG CAGCTCACTC G 321

(2) INFORMATION FOR SEQ ID NO:295

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 165 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

GACACACAGC GCCTCCGGCC CCGCACAGGG GGCATGTCCA GAGGTGCTGT GTGTCACCAA 60

CTGGTCTTCT AATTTGGAAG GAGTTGGAAA GGCCTTTTTG TTGATGAAAA GTTGGAAACA 120

GTGGCACATA TCTNAGAGGG AGGAACGAGG CAGCGTGGTG AAGCG 165

(2) INFORMATION FOR SEQ ID NO:296

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 315 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

CGAATACAGG TAGTGCCCAG CTGGTTGGGC TGGCCCAGGA AAATNCTGCT GTGTCAAATA 60

CTGCTGGCCA GGATGAAGCC ACAGCTAAGG CTGTGTTGGA GCCCATTCAG AGCACCAGTC 120

TAATTGGGAC TTTAACCAGG ACATCTGACA GTGAGGTTCC AGATGTGGAA TCTCGTGAAG 180

ATTTAATTAA AAATCACTAC ATGGCAAGNA TATTTGAATT TATTTCTCAG TTGCAGCTGG 240

CTGACAGTAA GTCAGTGCAT TTTTATGGCG ATT[illegible] ATTGTGTAAA AGACTNGGCT 300

TGGCTGNAAA GTCTA 315

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 244 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

AGTACGGTTN NCGCTNAAGC TTGATNATCG RATTGCCAAT CTNCATATTT GTGTTAGAAT 60

CATTTGTTTT TGTGTCTTCA TGTTTCTATA AGATAGGACC AATATTCTTT ATTGGGCTTT 120

GATTTTATTT TGTAACTTAA ATGTATTAAG GCAATAAATG TAATTTTCCA CTNAAAACTA 180

TCATTATAGA TTTGGTTACT ACCTACTGCT CAGCAATTTT TTTTCTTATC AAAATTCTTC 240

CTGG 244

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 152 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

CCTGAACAGG TAATGAGAAA AATTTACACA CAAGTGATTT TGAAAACAGA ATGGGTTGCT 60

TACAAATTAC AGGAAATGTT ATAACACAAA CCAGAAGAAT TCAATGGAAG GCAATAAGGG 120

ATTCTGAAAT GAAAATTATA AAAGTATCAN GA 152

(2) INFORMATION FOR SEQ ID NO:299:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

CGATGTTTTT AATGTCATCA CACGTTGTCT CAAAATGAGT GGTGGCATCA TATGTGCGGG 60

AAATAAAGAT CTGGCTTTCT GTTCCCAAGT CTTTTGGTAC CAGGAGGTCA CTGATGCTAA 120

CAAATTTCTG TTCAATTGGT TCCAAGAGCT CCAAAGCTGG TCTGATTTCC TTCTCAGGCT 180

CCTTGGTTTC CACAGTTGTA CTAACTATAG CAATGTACTT CCCTTGTGCT GCTACATTGT 240

GCGCAAAGGA GATCATGCAG ACGTAGATAT CTGACTTTCG ATTGACTTTG GTTCTGTGGA 300

ATAATGATCT GGCAGGAGTT GGCATCATTG GTGTTCTTTG ATGGGGGTGG CTGAGGGATG 360

CAAATAACCT CTTG 374

(2) INFORMATION FOR SEQ ID NO:300:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

GGCTCACCAA GCTCAGCAAG TACGTGTACT TCTTCGAGGC CTGCCGGCTG CTGCAGAAGA 60

TGATTGACAT CTCCCTGGAT GGCTTCCTGC TGACTCCGGT GCAGAAGATC TGCAAGTACC 120

CTCTGCAGCT GGCCGAGCTG CTCAAATACA CGCACCCCCA GCACAGGGAC TTCAAGGATG 180

TTGAAGCCGC CTTGCATGCC ATGAAGAACG TGGCCCAGCT CATCAACGAG CGGAAGGGTA 240

GACTTGAGAA CATCGACAAG ATTGCTCAGT GGCAGAGCTC CATAGAGGAC TGGGAGGGAG 300

AAGGATCTCT TGGTCAGGAG CTCAGAACTC ATCTACTCGG GGGGAGCTGA CCTCGGGTTA 360

CACAG 365

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

GGTATTCAAA CAAATAGCCT GAGAATTTNG GGGGGATCTG AAATAGAGTA CTATGCTATG 60

TTGGCTAAAA CTGGTGTCCA TCACTACAGT GGCAATANTA TTGAACTGGG CACAGCATGC 120

GGAAAATACT ACAGAGTGTG CACACTGGCT ATCATTGATC CAGGTGACTC TGACATCATT 180

AGAAGCATGC CAGANCAGAC TGGTGAAAAG TAAACCTTTT CACG 224

(2) INFORMATION FOR SEQ ID NO:302:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 363 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

AGTTTCACTC TTGTTGCCCA GGCTGGAGTG CAATGGCGTG ATCTCGGCTC ASTGCAATCN 60

GCACCTTCCG GNTTCAAGCG ATTCTCCTGC CTCAGCCTCC CAAGTAGTTG GGATTACAGG 120

CATGCCACAC CATGCCCGGC CAATTTTTNA TTTTTCGTAC AGACAGGGTT TCTCCATGTT 180

GGTCAGGCTG GTCTCAAACT CCGAACCTCG GTGATCCGCC CACCTCGGCC TCCCAAAGTG 240

CTGGGATTAT AGGCATGAGC CACTGTGTCC GGCCAGCCTC AACAATTTTA ATGTTTCTTT 300

CAAGNCTATT AGAAACCTTT AATTGCTTCT TAAGTTTCTC CCCCAACTAT GGAGGAAGCA 360

TAT 363

(2) INFORMATION FOR SEQ ID NO:303:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 253 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

ATGCAGGAAS ATCTACCARG CAAATCGAAA ACAAAAAAAG GCAGGGGTTG CAATCCATCT 60

CTCTGATAAA ACAGACTTTA AACCAACAAR RRTCAAAAGA CACAGAGARG GCCATARCAT 120

AATAGTAAAG CGGATCAATT CAACAAGAAG AGCTAACTAT CCTAAATATA TATGCACCCA 180

ATACAGGAGC AACTAGATTC ATAAAGCAAG TCCTGGAGGT GCCTACAGAG GAGGCTTAGG 240

CTCCCACACA TTA 253

(2) INFORMATION FOR SEQ ID NO:304:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

TTTTTTTGAG ATGGAGTACT CGCTCTCTTG CCCGGGCTGG AGTGCAGTGG CGCGATCTCG 60

GCTCACCTGC AACCCCTGCC TCCCCAGTTC AAGAGGTTCT CCTGCCTCAG CCTCCCGGGT 120

GGCTGGAATT GCAGGCACAC ACCACCATGC CCAGCTGCTT TCTTGTATTT TTAGTGGAGA 180

CGTGGTTTCA CCATGTTGGC CAGGCTGGTC TTGAGCTCCT GACCTTAAGT GATCCGCCAG 240

CCTTGGCCTC CCAAAGTGCT GGGATTACAG GCGTGAGCAC CGTGCCCAGG CTGTTTTTTA 300

ACTGACTTTG GATTTTACTC CCTTTCTATG CAAATTTATT TTAGAATCTG TTCCTTAACC 360

TTAGGGGGTT GGGTTAGACA AGTTTCAAGG GAGCCTCAAG TGKAAATTGC TTAAGG 416

(2) INFORMATION FOR SEQ ID NO:305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 223 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

CACACCCAGC TAATTTTTGT ATTTTTAGTA GAGACGGGGT TTCACCATGT TGGCTTGGCT 60

GGTCACGAAC TCCTGGCCTT GAGTGATCCC CCTGCCTCAG CCTCCCAAAG TGCTGGGATT 120

ACAGGTGTGA GTCAGCGTGC CCAGCCCAGA TTTTATTGTT TTAATTACAA ATTTTACGTA 180

AGTTGTTTCT GCACATTTAT ATTTGCACAC TTGTGCTAGT GAG 223

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

GTTTTGCCAC ATTGGCCAGG CTGGTCTCGA ACTCCCGACC WWGTGAGCCA CCTGCCTTGG 60

CCTCTCAAAG TGCTGGGATT ACAGGCGTGA GCACCACGCC CGACCCATAG CTCTTTACAA 120

CTGCCTTGTA AAGAAAGCAT CATTTGGCAC TGTTAGTATT TCTCTTGAA 169

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

GATTTGGTAC AGAGTATGTG AGGAAGACAA CTCAGATTGC CATTTTAAAT AAAGTTGTAC 60

ATGAAGAATA ATTGGAATCA TCAGGTAATT TTTTTAAACA AAGGTTCTTC ATTTACTGTT 120

ATGATTGGAA AAAAAATTAG AAAATAAAGT AAGTSCCATA GGCTAATTAA AAAATAAAAC 180

CTTGGCCGGG CGGGGTGGCT TACGCCTATA ATCCCAGCAC TTTGGGAGGC CGAGACGGGC 240

AGATCACGNG GTCAGGAGAT TGAGACCATC CTGGCTAACA CGGTGAAACC CCATCTCTAC 300

TTG 303

(2) INFORMATION FOR SEQ ID NO:308:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

ATCTAGGAGG CTGAGGTGGG ATCGCCCCAG TACTGGAGGT GAGGGCTGCA GTGAGCCATG 60

ATCATTGCCA CACACTCCAN CCTGGGTGAC AGAGTGAGAC CCTCTSTCAA AAAAGCTGAG 120

TGAATYGAAA GATAGAGTAT ATT 143

(2) INFORMATION FOR SEQ ID NO:309:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

CCCACCCTCA TAANCCCCAC TGGGGAGTCT GGGGGCCTCT ATTGCCATGT GCCTGGAATN 60
ATNATATGCT CATCACTTTA TGAAGAATAA AATTTGTNTT TCCTGCCTTA AAGTTACATT 120
CGTTCTTCCG CTCAAATCCT GATCTGGTCC ATTAAAGAGT GTTCGCAGAC AAAGTTTCTG 180
AAAGATTAGA GAAGAATCC 199

(2) INFORMATION FOR SEQ ID NO:310:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

TCCCTGTACC ACCTCTTCCT GAATACGGAG GAAAAGTTCG TTATGGACTG ATCCCTGAGG 60
AATTCTTCCA GTTTCTTTAT CCTAAAACTG GTGTAACAGG ACCCTATGTA CTCGGAACTG 120
GGCTTATCTT GTACGCTTTA TCCAAAGAAA TATATGTGAT TAGCGCAGAG ACCTTCACTG 180
CCCTATCAGT ACTAGGTGTA ATGGTCTATG GAATTAAAAA ATATGGTCCC TTTGTTGCAG 240
ACTTTGCTGA TAAACTCAAT GAGCAAAAAC TTGCCCAACT AGAAGAGGCG AAGAAGTTCT 300
TCCATCCAAC ACATCCAGAA TGCAATTGGA TACGGAGAAG GTCACAACAG GCACTGGTTT 360
CCAGGAAGCG CCATTTACCG TTTTTMATGG GMCAAAGGGA GTTACATTGG CTATGGCTTT 420
TGGAAG 426

(2) INFORMATION FOR SEQ ID NO:311:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

TCGACTCGGT CCTGGATGTG GTGAGGAAGG AGTCAGAGAG CTGTGACTGT TTCCAGGGCT 60
TCCAGCTGAC CCACTCTCTG GGGGGCGGCA CGGGGTCCGG GATGGGCACC CTGCTCATCA 120
GCAAGATCCG GGAAGAGTAC CCAGACCGCA TCATGAACAC CTTCAGCGTC ATGCCCTCAC 180
CCAAGGTGTC AGACACGGTR GTGGAGCCCT ACAACGCCAC CCTMTCGGTC CACCAGCTGG 240

TGGAAAACAC AGATGAAACC TACTGCATTG ACAACGAGGC CCTGTATGAC ATCTGCTTCC 300

GCACCCTGAA GCTGACCACC CCCACCTACG GGGACCTCAA CCACCTGGTG TCGGCCACCA 360

TGAGCGGGGT AACACCTGCT TGCGCTTYCC GGGCCAGCTG AACGAGACCT GGCAAAGTGG 420

CGGTTGACAT GGTGCCTTTC CTGGCTGAAT TTTTAATGCC CGGTTTGGGC CCTACCAGCC 480

GGGGAAGCA 489

(2) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

CTTCTCATGC CAGTCTAATG ATTGTTTTTA GAAAAGGATA TACATTGACC TTCAATGTAA 60

TAAGAAATGC AACACTTTAC GGTGTCCAAC TGCTAAGATT TATTTCCAAC TTGTCAGACA 120

CAACTATTTT GCCCAATCCA AATCAAAGGG AATCAAGGCT GTGAAATCCA CACAGGACAT 180

CAACGCACAC ATAAATGAAA ACTACAGATG TGTCAGAGGC AACCATATAC ACACAAATAA 240

TGTAACTACT AAATTCCATG AAGTAGCTGT CCAGGGAATA CTTTCCAAAT AACCTTCAGC 300

AG 302

(2) INFORMATION FOR SEQ ID NO:315:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

CGCGTTATTT AAATTGTGAA AAATAATGAA TATTAATTTG GAGCATAATA TTTAAATACA 60

TGAAAAAAGC TGGCTGCGAA ATGTTGGCAT GACTTTTCGC AGATGTTAGC ACTGCTTCAA 120

CTTTTGAGAG NGCACTCTGA GTGTAAGTTT ACTAGACTGA CATTACTAAA ATCATTGGTG 180

CTATAGAGGC AGGAGAATAC GGGGAATAAG AAAGCCAGTT GCAAGCCAAC AATCCTAAAA 240

CTGCTCCTTT TGGCATGGAC TGACGGCATA TTAAATGAGA TCATGCATTT TAAGGNATTA 300

AGAGTGTAGA CCACATGTGC GTGTTCCAAT AAAAGGAAG 339

Note regarding Claims: Certain SEQ ID NOS are excluded from some claims based on their homology to known non-human sequences (See Table 2).

WHAT IS CLAIMED IS:

1. An enriched oligonucleotide having a sequence designated as one of:

SEQ ID NO: 1 - 315;

or having a sequence complementary thereto.

2. An enriched oligonucleotide having a sequence designated as one of:

SEQ ID NO: 1 - 315, except SEQ ID NOS: 22 or 187;

or allelic variation or complementary sequence thereto or portion thereof at least 15 nucleotides in length.

3. An isolated oligonucleotide that includes a sequence designated as one of:

SEQ ID NO: 1 - 315, except SEQ ID NOS: 22, 187;

or allelic variation or complementary sequence thereto or portion thereof at least 15 nucleotides in length.

4. An enriched or isolated oligonucleotide operably coding for a human gene product, which includes a region coding for the same amino acid sequence as the coding region of a gene corresponding to a sequence designated as one of:

SEQ ID NO: 1 - 315.

5. The sequence of Claim 4, wherein said SEQ ID NO is listed in Table 6.

6. The sequence of Claim 4, wherein said SEQ ID NO is listed in Table 7.

7. The sequence of Claim 4, wherein said SEQ ID NO is identified in Table 10 in a metabolic functional grouping.

8. The sequence of Claim 4, wherein said SEQ ID NO is identified in Table 10 in a structural functional grouping.

9. The sequence of Claim 4, wherein said SEQ ID NO is identified in Table 11 in a developmental control grouping.

10. An enriched or isolated oligonucleotide coding for a human gene product, which includes a coding region

corresponding to the EST identified as:

SEQ ID NO: 1 - 315;

or a sequence complementary thereto or comprising an allelic variation thereof.

11. The oligonucleotide of Claim 10, wherein said SEQ ID NO is 1-315.

12. The oligonucleotide of Claim 10, wherein the SEQ ID NO is 1001-1500.

13. The oligonucleotide of Claim 10, wherein the SEQ ID NO is 1501-2000.

14. The oligonucleotide of Claim 10, wherein the SEQ ID NO is 2001-2421.

15. The oligonucleotide of Claim 10, wherein said sequence further includes the entire sequence designated as any one of SEQ ID NOS: 1-315.

16. An enriched or isolated oligonucleotide fragment comprising at least 15 bp of a sequence of Claim 10 and wherein said SEQ ID NO excludes NOS 22 and 187.

17. An enriched or isolated oligonucleotide sequence corresponding to a human gene, which hybridizes to a sequence designated as any one of SEQ ID NOS 1-315, except SEQ ID NOS 22, 187, or to a sequence complementary thereto, under hybridization conditions sufficiently stringent to require at least 97% base pairing.

18. An oligonucleotide according to any one of Claims 1-17, in substantially purified form.

19. A construct comprising a vector and an oligonucleotide according to any one of Claims 1-17.

20. The construct according to Claim 19, further comprising a promoter operably linked to said oligonucleotide.

21. A panel of at least 100 oligonucleotides according to Claim 3 or Claim 16.

22. An antisense oligonucleotide capable of blocking expression of the gene product of any one of the sequences

of Claim 10.

23. A triple helix probe capable of blocking expression of the gene product of any one of the sequences of Claim 10.

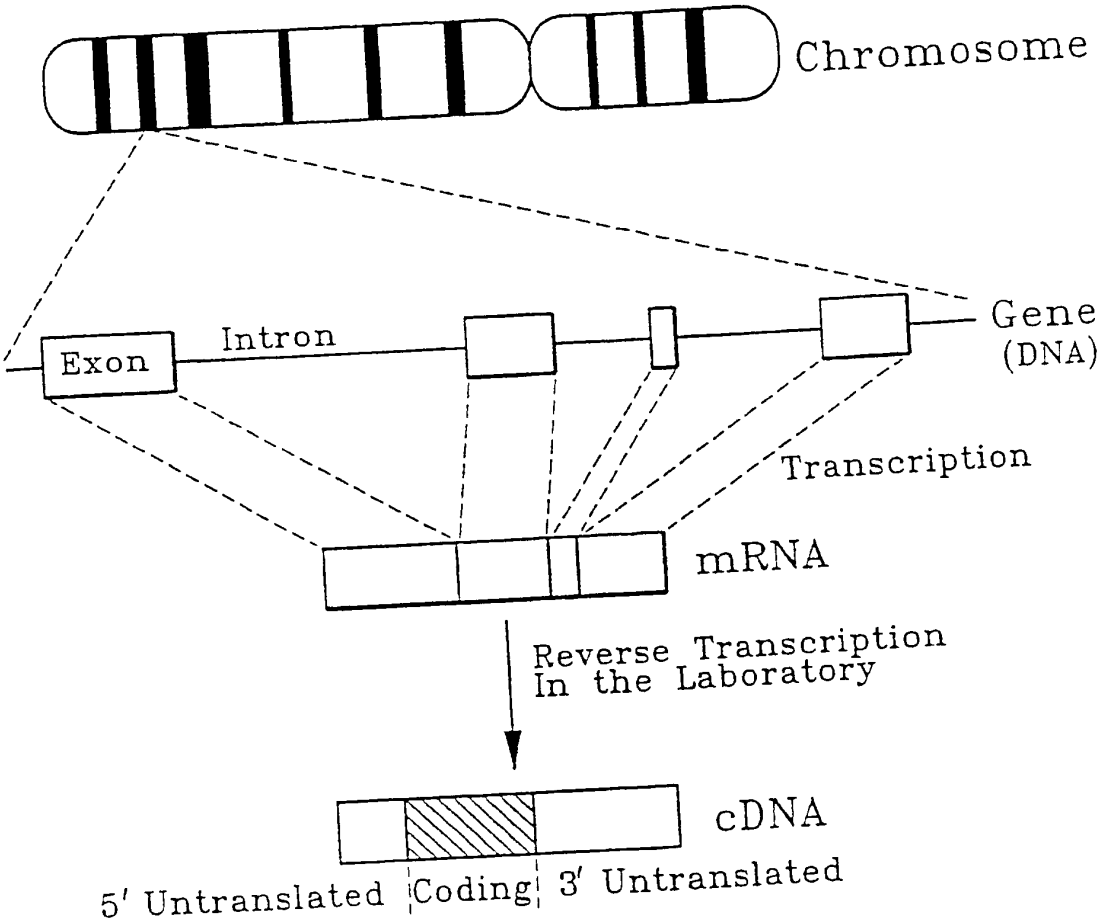
Chromosome
Gene
(DNA)
Exon
Intron
Transcription
mRNA
Reverse Transcription
In the Laboratory
cDNA
5′ Untranslated
Coding
3′ Untranslated

FIG. 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/05222

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07H 21/04, 21/02; C12N 15/11, 15/00
US CL :536/27; 435/320.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/27; 435/320.1, 6, 172.3, 91; 935/5, 6, 8, 9, 19, 22, 29, 80

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Please See Extra Sheet.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Proceedings of the National Academy of Sciences USA, Volume 80, issued January 1983, B.J. Conner et al, "Detection of sickle cell ßS-globin allele by hybridization with synthetic oligonucleotides", pages 278-282. See entire document.	1-11,15-23
X	Pharmacia P-L Biochemicals 1984 Product Reference Guide, published 1984 by Pharmacia P-L Biochemicals, Inc., Piscataway, NJ, USA, pages 36-37. See especially "Oligo(dA)" and "Oligo(dT)".	1-4, 10, 11, 15-18, 22, and 23
X	Cell, Volume 3, issued December 1974, P.C. Wensink et al, "A system for mapping DNA sequences in the chromosomes of Drosophila melanogaster", pages 315-325. See entire document.	19
X	Promega Biological Research Products 1988/89 Catalog, published 1988 by Promega Corporation, Madison, WI, USA. See entire document.	19 and 20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
08 September 1992

Date of mailing of the international search report
15 SEP 1992

Name and mailing address of the ISA/
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer
JAMES MARTINELL

Facsimile No. NOT APPLICABLE

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1992) *

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/05222

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Science, Volume 205, issued 1979, J.A. Martial et al, "Human growth hormone: Complementary DNA cloning and expression in bacteria", pages 602-606. See entire document.	21
Y	Gene, Volume 88, issued 08 June 1990, P. Szafranski et al, "Hypersensitive mung bean nuclease cleavage sites in Plasmodium knowlesi DNA", pages 141-147. See especially Figure 6 on page 145.	1-4, 10, 11, 15-20, 22, and 23
Y	Plant Molecular Biology, Volume 11, issued 1988, T.J. Higgins et al, "The sequence of a pea vicilin gene and its expression in transgenic tobacco plants", pages 683-695. See especially Figure 1 on page 686.	1-5, 10, 11, 15-20, 2, and 23
Y	Nature, Volume 338, No.17, issued 02 March 1989, G.H. Travis et al, "Identification of a photoreceptor-specific mRNA encoded by the gene responsible for retinal degeneration slow (rds)", pages 70-73. See especially Figure 3 on page 73.	1-4, 7, 10, 11, 15-20, 22, and 23
Y	The Journal of Biological Chemistry, Volume 264, No. 17, issued 15 June 1989, S. Matsuura et al, "Human adenylate kinase deficiency associated with hemolytic anemia", pages 10148-10155. See especially Figure 2 on page 10154.	1-4, 6, 10, 11, 15-20, 22, and 23
Y	The Journal of Biological Chemistry, Volume 263, No.6, issued 25 February 1988, S. Memet et al, "RPA190, the gene coding for the largest subunit of yeast RNA polymerase A", pages 2830-2839. See especially Figure 4 on page 2833.	1-4, 9-11, 15-20, 22, and 23
Y	Nucleic Acids Research, Volume 11, No. 12, issued 1983, Rosenzweig et al, "Sequence of the C. elegans transposable element Tc1", pages 4201-4209. See especially Figure 2 on page 4205.	1-4, 8, 10, 11, 15-20, 22, and 23
Y	Proceedings of the National Academy of Sciences USA, Volume 78, No. 11, issued Novermber 1981, S.V. Suggs et al, "Use of synthetic oligonucleotides as hybridization probes: Isolation of cloned cDNA sequences for human ß2-microglobulin", pages 6613-6617. See entire document.	1-11, 15-20, 22, and 23
Y	Analytical Biochemistry, Volume 172, issued 1988, C.J. Marcus-Sekura, "Techniques for using antisense oligodeoxyribonucleotides to study gene expression", pages 289-295. See entire document.	1-11, 15-20, 22, and 23
Y	Methods in Enzymology, Volume 152, issued 1987, A.R. Kimmel, "Selection of clones from libraries: Overview", pages 393-399. See entire document.	1-11, 15-20, 22, and 23
Y	Methods in Enzymology, Volume 101, issued 1983, M. Rosenberg et al, "The use of pKC30 and its derivatives for controlled expression of genes", pages 123-138. See entire document.	1-11, 15-20, 22, and 23
Y	Nucleic Acids Research, Volume 12, No. 18, issued 1984, D.A. Melton et al, "Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter", pages 7035-7056. See entire document.	1-11, 15-20, 22, and 23
Y	Proceedings of the National Academy of Sciences USA, Volume 83, issued 1986, A. Hirashima et al, "Engineering of the mRNA-interfering complementary RNA immune system against viral infection", pages 7726-7730. See entire document.	1-11, 15-20, 22, and 23

B. FIELDS SEARCHED
Documentation other than minimum documentation that are included in the fields searched:

USB MOLECULAR BIOLOGY REAGENTS/PROTOCOLS 1992, UNITED STATES BIOCHEMICAL
GENES, LEWIN, 1992, JOHN WILEY & SONS, NEW YORK, NY
PHARMACIA P-L BIOCHEMICALS 1984 PRODUCT REFERENCE GUIDE
PROMEGA BIOLOGICAL RESEARCH PRODUCTS 1988-89 CATALOG
STRATAGENE CLONING SYSTEMS 1992 PRODUCT CATALOG

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

CAS ONLINE, APS, IGSUITE
Search Terms: expressed sequence tag?, est, ests, gene#, dna#, cdna#, rna#, mrna#, librar?, brain?, hippocampus, temporal cortex
SEQ ID NO 1 15-mers in nucleotide positions no 1-100, SEQ ID NO 7 15-mers in nucleotide positions 1-34, SEQ ID NO 9 15-mers in nucleotide positions 1-54, SEQ ID NO 20 15-mers in nucleotide positions 1-54, SEQ ID NO 43 15-mers in nucleotide positions 1-25, and SEQ ID NO 77 15-mers in nucleotide positions 1-43.

A NOTE ON THE SEARCH

An exhaustive search of all of the oligonucleotides embraced by the claims has not been undertaken. It is noted that a very large number of oligonucleotides is in fact embraced by the claims. The instant application discloses 308 sequences that contain a total of just over 100,000 nucleotides. The total number of 15-mers contained in these sequences is about 100,000. Thus, an exhaustive search of 15-mers only would require 200,000 searches; 100,000 searches of the 15-mers plus 100,000 searches of the complements of each of the 15-mers. One should note that this large number of searches does not include oligonucleotides longer than 15 nucleotides, nor does it include a consideration of sequences that do not precisely match the sequences that are disclosed (e.g., "allelic variations"). Nor does this take into account any errors in sequencing that are mentioned at page 17, lines 26-33 of the instant application. The nucleotide sequence searching equipment at the USPTO is capable of searching a 15-mer across a collection of databases that includes over 100,000,000 nucleotides in about 15 minutes. Even at this high rate of speed, a complete and exhaustive search of all of the 15-mers that are embraced by the claims could not be completed before January 1998. Therefore, a search designed to determine whether the claims presented were novel or would not require an inventive step was performed.

This is how the search was done. The claims were inspected to determine the minimum number of separate sequences that would need to be searched in order to find a 15-mer contained within a gene in the database and included in each of claims 1-11 and 15-23. The following sequences (i.e. SEQ ID NOs) were selected to represent the claims: 1, 7, 9, 20, 43, and 77. The sequence databases were then selected for 15-mers in the following manner. The first 15 (i.e. positions 1-15 inclusive) nucleotides of a sequence were used as a search query against the sequence databases. Then, the second 15-mer (i.e. positions 2-16) was searched, and so on, crawling down the sequence until a match for which a reference published prior to 1990 was found in the database. Because the searching of a given sequence was performed only until a reference was found, an equal number of searches for each SEQ ID NO was not necessary. The following table shows the number of searches completed.

SEQ ID NO	NUMBER OF 15-MERS SEARCHED
1	86
7	20
9	40
20	40
43	11
77	19